Epidemiological and genomic landscape of antimicrobial resistance in Malawi

A thesis submitted in accordance with the requirements of the

University of Liverpool for the degree of Doctor of Philosophy

Patrick Musicha

August 2017

University of Liverpool

DECLARATION

This work has not been previously accepted in substance for any other degree and is not concurrently being submitted in candidature for another degree.

This thesis is the result of my own work. Work done in collaboration with others is acknowledged and detailed below.

ACTIVITY	RESPONSIBILITY
Blood culture and CSF isolates collection	QECH and MLW clinical and Diagnostic Laboratory Staff
Rectal swab isolates	Margaret Khonga
Microbiology procedures	MLW Diagnostic Laboratory staff
DNA Extraction	Chikondi Peno
Library preparation, whole genome sequencing and pre-processing	Wellcome Trust Sanger Institute Core Sequencing Team

LIST OF UNPUBLISHED SCRIPTS USED IN THIS THESIS

SCRIPT NAME	USE	Author
iCANDY.py	Phylogenetic tree annotation	Simon Harris
findKclusters.py	Polysaccharide capsule (K) serotyping of *K. pneumoniae* isolates	Rainer Follador
Get_sequence_type.py	*In silico* MLST	Simon Harris
Map_resistome.py	Pipeline for BLAST searching against AMR genes	Simon Harris
SMALT	Mapping/aligning read sequences to reference genome sequence	Hans Ponstingl
Primers_to_tab.py	*In silico* PCR	Simon Harris

Signed……

PATRICK MUSICHA

Funding: This PhD was funded by the National Institutes for Health of the United States through the H3Africa Bioinformatics Network (H3ABionet) Project

ABSTRACT

Antimicrobial resistance (AMR) is a global public health problem, which presents a huge threat to the treatment of all forms of bacterial infections. A wide range of bacterial pathogens across the globe are increasingly developing resistance to multiple classes of antimicrobial agents rendering the agents concerned ineffective for the treatment of infections. Bloodstream infection (BSI) and other bacterial infections in sub-Saharan Africa (SSA) and Malawi in particular, are a common cause of morbidity and mortality.

Few facilities in SSA however, are able to conduct long-term surveillance and as such the full burden of drug resistant infection (DRI) remain largely unknown across the region. In this thesis, blood cultures routinely taken from adult and paediatric medical patients admitted to Queen Elizabeth Central Hospital (QECH) in Blantyre, Malawi between 1998 and 2016 were analysed to describe trends in BSI and AMR.

The analysis revealed a significant decline of BSI in all major pathogens except *S.* Typhi. However, the majority of isolates were resistant to the Malawian first-line antimicrobial agents (ampicillin, cotrimoxazole and chloramphenicol). Resistance to all the first line antimicrobial agents was more common in Gram-negative pathogens than Gram-positive pathogens. Non-*Salmonellae* Enterobacteriaceae that produced extended spectrum beta-lactamase (ESBL) and were fluoroquinolone-resistant were detected, and the proportions of these isolates rose significantly during the

surveillance. In contrast, a majority of common Gram-positive pathogens remain susceptible to either penicillin or chloramphenicol. Methicillin resistant *S. aureus* was first reported in 1998 but became regularly detected in the later years of the surveillance.

The analysis of blood culture isolates identified *E. coli* as one of the common causes of BSI in Blantyre, and the proportion of these isolates that were ESBL producers increased over time. Globally, efforts to treat *E. coli* infections are increasingly being compromised by the rapid, global spread of ESBL-producing *E. coli*. In this thesis, a whole genome sequencing (WGS) study was carried out to investigate the genetic population structure and molecular determinants of AMR in *E. coli* isolates from Malawi.

Whole genomes of clinical *E. coli* isolates from patients admitted to QECH were sequenced and analysed using phylogenetic methods and comparative genomics. It was revealed that the *E. coli* population in Malawi is highly diverse with isolates belonging to five phylogroups, corresponding to five isolate sequence clusters (SCs) that contained over forty sequence types (STs). A unique sub-lineage of ST131 was identified that was distinct from previously defined sub-lineages of this globally disseminated ST. The most common ESBL gene was $bla_{\text{CT X-M-15}}$. Unlike in other settings where presence of the $bla_{\text{CT X-M-15}}$ gene was strongly linked to ST131, here the gene was not lineage-specific suggesting a distinct genomic landscape of ESBL-producing *E. coli* in Malawi.

This thesis also identified *Klebsiella spp.* isolates as a common cause of BSI in Blantyre, and an increasing proportion of ESBL-producing and fluoroquinolone resistant isolates were identified. The molecular mechanisms and clones of *K. pneumoniae* associated with ESBL production and fluoroquinolone resistance were yet to be explored in Malawi. Here, a number of *K. pneumoniae* isolates were selected for WGS, and placed in a global context by comparison with previously sequenced *K. pneumoniae* isolates from multiple locations outside SSA, in order to identify the molecular determinants of AMR and determine their relationship with *K. pneumoniae* population structure.

Genomic analysis revealed three main lineages of *K. pneumoniae*, which corresponded to the previously defined KpI, KpII and KpIII lineages. All three lineages exhibited high genetic diversity. Further phylogenetic analysis revealed a sub-lineage of KpI to be a major cause of CA infections in Malawi. The sub-lineage included the clonally related ST14 and ST15 of *K. pneumoniae* which cause hospital acquired infection in multiple settings across the globe, A large pool of AMR genes, was identified in the genomes of the Malawian isolates, including multiple ESBL and *qnr* genes. Plasmid-encoded CTX-M-15 was the most common type of ESBL that was identified. In common with *E. coli* from Malawi, AMR was not restricted to a particular clade of *K. pneumoniae*. These findings suggest that dissemination of AMR in the *K. pneumoniae* population in Malawi was either due to a combination of horizontal gene transfer and clonal expansion, or horizontal gene transfer alone.

In conclusion, the thesis has shown that ESBL production and fluoroquinolone resistance is rapidly spreading in Malawi across multiple *E. coli* and *K. pneumoniae* lineages that are causing increasing levels of infection. As cephalosporins and fluoroquinolones remain the last resort antimicrobial agents in this setting, urgent action is needed to curb the spread of Gram-negative AMR pathogens.

DEDICATION

Mwanjiwa, Nova and Mahara,

This is probably not the greatest of achievements; but it will be great for me if it inspires you to achieve that which is greater for you.

ACKNOWLEDGEMENTS

My undertaking of this PhD relied on the support and encouragement that I received from many people, without whom progress of this PhD would have been very difficult. Firstly, I would like to thank my wife Mwanjiwa for enduring my absences from home (including during the birth of Mahara) and taking a larger share of our parental roles and keeping on nagging me to not ignore my other important responsibilities.

My sincere thanks to my supervisors Drs Nicholas Feasey, Chisomo Msefula and Dean Everett for believing in me and supporting me on my academic and personal life. You guided me well whilst giving me the freedom to advance my own scientific interests. My thanks also go to Profs. Robert Heyderman and Moffat Nyirenda who were part of the supervisory team at the beginning.

I am also very grateful to Prof. Nick Thomson for sequencing of the isolates used in this thesis and hosting me at the Sanger Institute. I thank Dr. Gabriela Gomes for her support with the mathematical modelling work, and Prof. Oleg Reiva for hosting me as a bioinformatics intern at the University of Pretoria. My Advisers, Prof. Melita Gordon and Dr. Wilson Mandala, for their advice that kept me on the right track.

I am indebted to many people I have met and worked with in the various institutions I worked during the PhD. In particular, Chrispin Chaguza (your friendship and of course, our endless arguments are among the best things I gained from joining MLW), Jen Cornick, Anmol Kiran, and all members of the MiDep Group

at MLW. Khuzwayo Jere, Yolanda and Benjamin Kumwenda, for the company in Liverpool. Amy Cain, Alison Mather and Teemu Kallonen, for ensuring that I was comfortable at WTSI.

Above all, I thank God for all that I am.

PUBLICATIONS AND PRESENTATIONS

Peer Musicha P, Cornick, JE, Bar-Zeev N, French N, Masesa C Denis B, Kennedy N, Mallewa J, Gordon MA Msefula CL, Heyderman RS, Everett, DB, Feasey NA. *Trends in antimicrobial resistance in bloodstream infection isolates at a large urban hospital in Malawi (1998-2016): a surveillance study.* **Lancet Infectious Diseases**. 2017 Aug 14. DOI: http://dx.doi.org/10.1016/S1473-3099(17)30394-8.

Musicha P**,** Feasey NA, Cain AK, Kallonen T, Chaguza C, Peno C, Khonga M, Thompson S, Gray KJ, Mather AE, Heyderman RS, Everett DB, Thomson NR, Msefula CL. *Genomic landscape of extended-spectrum β-lactamase resistance in Escherichia coli from an urban African setting*. **Journal Antimicrobial Chemotherapy** 2017

Other publications during the PhD

Chaguza C, Cornick JE, Andam C, Gladstone R, Alaerts M, Musicha P**,** Peno C, Bar-Zeev N, Kamng'ona A, Kiran A, Msefula CL, McGee L, Breiman R, Kadioglu A, French N, Heyderman RS, Hanage B, Bentley S, Everett DB. *Population genomic epidemiology of invasive Streptococcus pneumoniae isolates before introduction of the PCV13 vaccine in Malawi, 2004-2010*. **Vaccine** 2017.

Mulder NJ, *et al* (As H3ABionet Consortium). *H3ABioNet, a sustainable pan African bioinformatics network for human heredity and health in Africa*. **Genome Research** 2016.

Oral Presentations

Genomic landscape of extended spectrum β-lactamase resistance in *E. coli* isolates from an urban African setting, 2016-Wellcome Trust Liverpool Glasgow Centre for Global Health Research Annual Scientific Meeting, Chester, United Kingdom (Invited plenary talk), November 2016.

Posters

Genomic landscape of extended spectrum β-lactamase resistance in *E. coli* isolates from an urban African setting, 2016- Faculty of Health and Life Sciences poster day, University of Liverpool, June 2016.

Emergence of ESBL and fluoroquinolone resistance in Blantyre, Malawi-Liverpool-Wellcome Trust Clinical Research Programme Annual Scientific Meeting, Mangochi, Malawi, September 2015.

TABLE OF CONTENTS

DECLARATION ..... **i**

ABSTRACT ..... **ii**

DEDICATION ..... **vi**

ACKNOWLEDGEMENTS ..... **vii**

PUBLICATIONS AND PRESENTATIONS ..... **ix**

TABLE OF FIGURES ..... **xix**

TABLE OF TABLES ..... **xxiii**

TABLE OF APPENDICES ..... **xxv**

LIST OF ABBREVIATIONS ..... **xxvi**

Chapter one: General Introduction **1**

1.1 Background **1**

1.2 A brief history of antimicrobial agents and antimicrobial resistance in bacteria **2**

1.3 Mechanisms of action of antimicrobial agents **5**

1.3.1 Cell wall synthesis inhibitors 7

1.3.2 Protein synthesis inhibitors 8

1.3.3 Nucleic acid synthesis inhibitors 9

1.4 Antimicrobial agents used for treatment of SBI in Malawi **10**

1.5 Multidrug resistance **11**

1.6 Molecular mechanisms of AMR to agents in use in Malawi **12**

1.6.1 Acquisition of AMR 12
1.6.2 Mechanism of β-lactamase resistance 14
1.6.3 Mechanisms of fluoroquinolone resistance 17
1.6.4 Mechanism of resistance to aminoglycosides 19
1.6.5 Mechanism of resistance to chloramphenicol 19
1.6.6 Mechanism of resistance to cotrimoxazole 20
1.7 AMR in major bacterial pathogens in sub-Saharan Africa 20
1.7.1 Severe bacterial infections in SSA 20
1.7.2 AMR in *Salmonellae* 23
1.7.3 AMR in *Escherichia coli* 26
1.7.4 AMR in *Klebsiella pneumoniae* 29
1.7.5 AMR in *Streptococcus pneumoniae* 32
1.7.6 AMR in *Staphylococcus aureus* 33
1.8 Molecular epidemiology of ESBL and fluoroquinolone resistance in SSA 35
1.8.1 Molecular determinants of ESBL resistance 35
1.8.2 Molecular determinants of fluoroquinolone resistance 37
1.8.3 ESBL producing and fluoroquinolone resistant clones of *E. coli* and *K. pneumoniae* 39
1.9 Factors driving the spread of AMR 42
1.10 Concluding remarks and rationale 44
1.11 Hypothesis and Aims 46
1.11.1 Hypothesis 46
1.11.2 Aims 47
Chapter Two: Methods and Materials 48

2.1 Overview **48**

2.2 Retrospective study of community acquired bloodstream infection and antimicrobial resistance in Blantyre **48**

2.2.1 Study setting 48

2.2.2 Microbiology procedures 51

2.2.3 Antimicrobial susceptibility testing 52

2.2.4 Data management and statistical analyses of surveillance 54

2.3 Whole genome sequencing of *E. coli* and *K. pneumoniae* isolates **59**

2.3.1 Study isolates 59

2.3.2 Processing of genomic DNA 61

2.3.2.1 Culturing of selected E. coli and K. pneumoniae isolates *61*

2.3.3 Genomic DNA sequencing 62

2.4 Bioinformatics analyses **62**

2.4.1 Quality checking and control for sequence data 62

2.4.2 Mapping of genome read sequences to reference genomes sequences 64

2.4.3 Genome sequence assembly and annotation 65

2.4.4 *In silico* multi-locus typing 65

2.4.5 *In silico* determination of *E. coli* phylogroups 67

2.4.6 Pan genome and core genome analyses 69

2.4.7 Inference of population structure 70

2.4.8 Phylogeny reconstruction 71

2.4.9 Identification of AMR molecular determinants 73

2.4.10 Plasmid replicon typing 73

2.4.11 Statistical analyses 74

2.5 Mathematical modelling of force of *E. coli* BSI **76**

Chapter Three: Trends in community acquired bloodstream infection and antimicrobial resistance in Blantyre 77

3.1. Overview 77

3.2. Introduction 78

3.3. Trends in CA-BSI in Blantyre 80

3.3.1. General overview of blood cultures from QECH, 1998-2016 80

3.3.2. Causes of BSI in Blantyre 81

3.3.3. Trends in incidence of BSI 87

3.3.4. Age distribution of patients with confirmed CA-BSI 89

3.3.5. Contaminants 93

3.4. Trends in Antimicrobial resistance 94

3.4.1. Partial return of chloramphenicol susceptibility in Enterobacteriaceae 96

3.4.2. Emergence of ESBL producing, fluoroquinolone and aminoglycoside resistant isolates 96

3.4.3. Antimicrobial resistance in Gram-positive pathogens 98

3.4.4. Multi-drug resistance 100

3.5. Discussion 105

3.6 Limitations 108

3.7 Conclusion 109

Chapter four: Genomic landscape of extended antimicrobial resistance in *Escherichia coli* from Blantyre, Malawi 110

4.1 Overview 110

4.2 Introduction 111

4.3 Sequenced Malawian *E. coli* isolates 112

4.4 Genetic diversity and population structure of Malawian *E. coli* isolates 115

4.4.1 Phylogroups and multi-locus sequence type diversity 115

4.4.2 Pan and core genomes 120

4.4.3 Population structure of Malawian *E. coli* 124

4.4.4 Malawian ST131 isolates in context of global population structure of ST131 129

4.5 Genetic determinants of antimicrobial resistance in *E. coli* genomes 131

4.5.1 Presence of AMR genes and phylogenetic clustering 132

4.5.2 Genetic determinants of β-Lactam resistance 133

4.5.3 Molecular determinants fluoroquinolone resistance 136

4.5.4 Molecular determinants of aminoglycoside resistance 140

4.5.5 Genetic determinants of chloramphenicol resistance 140

4.5.6 Molecular determinants of cotrimoxazole resistance 141

4.5.7 Plasmid incompatibility groups 141

4.6 Discussion 142

4.7 Limitations 144

4.8 Conclusion 145

Chapter Five: Population structure, genetic diversity and antimicrobial resistance of *K. pneumoniae* isolates from Malawi146

5.1 Overview 146

5.2 Introduction 147

5.3 Sequenced Malawian *K. pneumoniae* isolates 150

5.4 Genetic diversity of Malawian *K. pneumoniae* isolates 152

5.4.1 Multi-locus sequence types and polysaccharide capsule types of Malawian *K. pneumoniae* 152

5.4.2 Pan genome and core genome of Malawian *K. pneumoniae* isolates 154

5.4.3 The Malawian *K. pneumoniae* population has an open pan genome 156

5.5 Population structure of Malawian *K. pneumoniae* isolates 159

5.5.1 Bayesian population structure of Malawi *K. pneumoniae* 159

5.5.2 Reconstructed phylogeny of Malawian *K. pneumoniae* isolates 160

5.5.3 Malawian *K. pneumoniae* isolates in context of global *K. pneumoniae* population structure 162

5.5.4 Diversity of Malawian and global *K. pneumoniae* accessory genomes 164

5.5.5 Population structure of Malawian KpI isolates: Emergence and evolution of the ST14 and ST15 lineage 167

5.6 Genetic determinants of antimicrobial resistance 175

5.6.1 Acquired AMR genes in genomes of Malawian *K. pneumoniae* 175

5.6.2 Distribution of acquired AMR genes across the phylogeny of the Malawian *K. pneumoniae* 177

5.6.3 Molecular determinants of β-Lactam resistance 181

5.6.4 Molecular determinants of fluoroquinolone resistance 183

5.6.5 Molecular determinants of aminoglycoside resistance 184

5.6.6 Molecular determinants of chloramphenicol resistance 185

5.6.7 Molecular determinants of cotrimoxazole resistance 185

5.6.8 Plasmid incompatibility types associated with AMR 186

5.7 Discussion 188

5.8 Limitations 194

5.9 Conclusion 194

Chapter Six: A mathematical model estimating incidence of *E. coli* *BSI* in a Blantyre population with heterogeneous susceptibility 196

6.1 Overview 196

6.2 Introduction 196

6.3 Population with homogeneous susceptibility to infection 197

6.4 Population with discrete groups with different susceptibility levels 198

6.5 A heterogeneous susceptibility model provides a better fit to observed incidence better 200

6.6 Discussion 207

6.7 Limitations 208

6.8 Conclusion 209

Chapter seven: Concluding remarks and future direction 210

7.1 Overview 210

7.2 Summary of Key findings 210

7.2.1 Decline of bloodstream infection in Blantyre 210

7.2.2 Emergence and expansion of ESBL and fluoroquinolone resistant Enterobacteriaceae 211

7.2.3 The changing molecular epidemiology of ESBL resistance 212

7.2.4 Multiple *E. coli* and *K. pneumoniae* lineages carry genotypes encoding ESBL and fluoroquinolone resistance 213

7.2.5 There is underlying heterogeneity in susceptibility to bloodstream infection in Blantyre 215

7.3 Future work 215

7.4 Conclusions 217

APPENDICES..... 218

REFERENCES 258

TABLE OF FIGURES

Figure 1.1 Timelines in the discovery/development of antimicrobial agents and emergence of resistance. 3

Figure 1.2 An illustration of Antimicrobial (mechanisms of action that lead to bacterial cell death. 6

Figure 1.3 Cell walls of Gram-negative and Gram-positive bacteria 7

Figure 1.4 An Illustration of mechanisms of acquiring AMR genes and MGEs.. 13

Figure 1.5 An illustration of ESBL hydrolysis of the β-lactam ring. 16

Figure 1.6 Phylogenetic relationships of CTX-M-15 associated *E. coli* ST131 clone in context of other ST131 strains.. 41

Figure 1.7 World Map showing the relationship between travel routes and emergence of carbapenem, cephalospoirn and Linezolid resistance in Gram-negative bacteria, *N. gonorrhoea* and *Enterococci* respectively 44

Figure 2.1 Map of Africa and Malawi (Central Intelligence Agancy, 2017) 49

Figure 3.1 Distribution of surveillance blood cultures at QECH from 1998-2016. 80

Figure 3.2 Distribution of Enterobacteriaceae isolates from 1998-2016. 82

Figure 3.3 Distribution of BSI cases by non-Enterobacteriaceae Gram negative pathogens in Blantyre, 1998-2016. 85

Figure 3.4 Distribution of BSI cases by Gram positive pathogens and yeast in Blantyre, 1998-2016 86

Figure 3.5 Trends in incidence of BSI 1998-2016 88

Figure 3.6 Distribution isolates by age of patients with BSI 90

Figure 3.7 Estimated minimum incidence rates of BSI stratified by age..... 92
Figure 3.8 Distributions of contaminants by organism, expressed as proportions of all blood cultures per year of isolation. 93
Figure 3.9 Trends in proportions of isolates resistant to Malawian first line antimicrobials..... 95
Figure 3.10 Trends in resistance to second-line antimicrobial agents in *E. coli*, *Klebsiella spp*. and other Enterobacteriaceae 98
Figure 3.11 Trends in resistance to first-line agents in Gram-positive pathogens.... 99
Figure 3.12 Distribution of isolates, by year and organism, that had expressed co-resistance to ampicillin, chloramphenicol, cotrimoxazole, ceftriaxone, ciprofloxacin and gentamicin between 2003 and 2016 in Blantyre.....104
Figure 4.1 Distribution of *E. coli* isolates selected for WGS by year of isolation.....114
Figure 4.2 Determination of *E. coli* phylogroups 116
Figure 4.3 Distributions of genes in the *E. coli* isolates collected from Blantyre.....121
Figure 4.4 Expansion of the Malawian *E. coli* pan genome with increasing number of genome sequences as modelled by the power law-regression equation. 123
Figure 4.5 Distribution of isolates and pairwise SNP differences by SC.....125
Figure 4.6 Tanglegram comparing ML phylogeny based on core genome alignment with ML phylogeny based on whole genome sequence SNP alignment obtained from mapping to a reference genome. 127
Figure 4.7 Population structure Malawian *E. coli* isolates collected from Blantyre, Malawi.....128
Figure 4.8 Phylogeny of Malawian ST131 isolates in a global context.....130

Figure 4.9 Distribution of AMR phenotype profiles, acquired AMR genes and plasmid incompatibility groups across the phylogeny of Malawian *E. coli* isolates. 133
Figure 4.10 Illustration of genetic environment of main ESBL and fluoroquinolone resistance genotypes in Malawian *E. coli* isolates. 136
Figure 5.1 Distribution of isolates selected for WGS by clinical source of isolation. 150
Figure 5.2 A flow-chart showing the exclusion and inclusion criteria of *K. pneumoniae* genome sequences for further analyses after quality control checks 152
Figure 5.3 Distributions of genes in per given number of genome sequences of *K. pneumoniae* isolates from Blantyre, Malawi 156
Figure 5.4 Expansion of the pan genome size with increasing number of genome sequences as modelled by the power law-regression equation 158
Figure 5.5 Population structure of Malawian *K. pneumoniae* isolates. 161
Figure 5.6 Phylogroups of Malawian *K. pneumoniae* isolates 163
Figure 5.7 PCA Analysis of *K. pneumoniae* genomes based on accessory genes present in 5.0% - 95.0%of isolates 165
Figure 5.8 Venn diagram illustration of number of genes in the core genomes of 66 Malawian and 22 global KpI isolates. 167
Figure 5.9 Recombination sites in genomes of Malawian KPI isolates 168
Figure 5.10 Recombination sites in genomes of isolates in KpI-SC1 sub-lineage 169
Figure 5.11 Estimation of root-to-tip divergence over sampling time of Malawian *K-pneumoniae* isolates belonging to the KpI sub-cluster KpI-SC1. 172

Figure 5.12 Dated Bayesian phylogenetic tree of KpI-SC1 comprising of ST14 and ST15 isolates. 174
Figure 5.13 Distribution of AMR phenotypic profiles and acquired AMR genes across the phylogeny of Malawian *K. pneumoniae* isolates..... 178
Figure 5.14 Distribution of number of AMR genes per genome 180
Figure 5.15 Distribution of ESBL and fluoroquinolone resistance genotypes across the phylogeny of the Malawian *K. pneumoniae* isolates. 182
Figure 5.16 A heatmap illustration of the association between plasmid Inc-types and acquired AMR genes present in ≥ 5 genomes..... 187
Figure 6.1 Incidence of BSI in children aged < 5 years as predicted by the age-dependent models assuming a population with homogeneous susceptibility to BSI and a heterogeneous population with two groups with distinct susceptibility levels fitted on a bar plot of observed incidence of BSI.. 200
Figure 6.2 A diagrammatic illustration of the heterogeneous model over time 202
Figure 6.3 Incidence of *E. coli* BSI in children and adults in Blantyre over time estimated by the heterogeneous model fitted against incidence calculated from observed BSI cases. 206

TABLE OF TABLES

Table 2.1A Age stratified population denominators for Blantyre City (1998-2007) 56

Table 2.1B Age stratified population denominators for Blantyre (2008-2014) 57

Table 2.2 Seven house-keeping genes for *E. coli* MLST 66

Table 2.3 Seven housekeeping genes for Klebsiella MLST scheme..... 67

Table 2.4 Primer sequences for *E. coli* phylogroup typing 68

Table 2.5 Classification of *E. coli* isolates into phylogroups based on the presence or absence of gene prime sequences..... 69

Table 3.1 Prevalence of significant pathogens in four time intervals: 1998-2001, 2002-2005, 2006-2009 and 2010-2013, 2014-2016..... 84

Table 3.2A Frequency of multidrug resistance (MDR) in Gram-negative pathogens causing BSI in Blantyre between 1998 and 2016. 102

Table 3.2B Frequency of multidrug resistance (MDR) in Gram-positive pathogens causing BSI in Blantyre between 1998 and 2014 103

Table 4.1 Distribution of *E. coli* isolates selected for WGS by clinical source of

Table 4.2 Distribution of Malawian *E. coli* isolates by ST.....118

Table 4.3 Distribution of AMR genes among Malawian *E. coli* isolates 131

Table 4.4 Characteristics of CTX-M-15 associated *E. coli* isolates 135

Table 4.5 Characteristics of isolates with fluoroquinolone resistance genotypes*.139

Table 5.1 Distribution of STs in Malawian *K. pneumoniae* isolates.....153

Table 5.3 Recombination statistics of isolates in KpI-SC1 sub-lineage.....170

Table 5.4 Distribution of AMR genes present in at least five isolates.....176

Table 6.1 Summary of model parameters.....205

TABLE OF APPENDICES

Appendix 1: Classification of isolates into subgroups 218

Appendix 2A Number of blood cultures by subgroup, 1998-2007 222

Appendix 2B- Number of blood cultures by subgroup, 2008-2016 223

Appendix 3 List of *E. coli* isolates sequenced, closest species genome match, and assembly statistics 224

Appendix 4 Sequenced *E. coli* isolates metadata 227

Appendix 5 List of *K. pneumoniae* isolates sequenced, closest species genome matches and assembly statistics 230

Appendix 6 Metadata for sequenced *K. pneumoniae* isolates 233

Appendix 7 List of all AMR genes identified in Malawian *K. pneumoniae* genomes 235

Appendix 8 Genes identified in Gubbins predicted recombinant regions of Malawian *K. pneumoniae* KpI genome sequences 237

LIST OF ABBREVIATIONS

ABBREVIATION	Full name
ºC	Degrees celsius
%	Percent
AIDS	Acquired immunodeficiency syndrome
AMR	Antimicrobial resistance
ART	Antiretroviral therapy
BAPS	Bayesian analysis of population structure
BEAST	Bayesian evolutionary analysis by sampling trees
BSI	Bloodstream infection
CA	Community acquired
CO2	Carbon dioxide
CSF	Cerebrospinal fluid
DCS	Decreased ciprofloxacin susceptibility
DNA	Deoxyribonucleic acid
DRI	Drug resistant infection
ENA	European nucleotide archive
ESBL	Extended-spectrum beta-lactamases
HA-BSI	Hospital acquired bloodstream infection
HIV	Human immunodeficiency virus
ID	Identification
LB	Luria broth
LIMS	Laboratory information management system
LSTM	Liverpool School of Tropical Medicine
MCMC	Markov chain Monte Carlo
MDR	Multidrug resistant
MGE(s)	Mobile genetic elements
MIC	Minimum inhibitory concentration
ML	Maximum likelihood
MLW	Malawi-Liverpool-Wellcome Trust Clinical Research Programme
MRCA	Most recent common ancestor
mRNA	Messenger rna
MRSA	Methicillin resistant staphylococcus aureus
NSO	National statistical office (Malawi)
NTS	Non-typhoid salmonella
p	p-value
PBP	Penicillin binding proteins

PCA	Principal component analysis
PCV	Pneumococcal conjugate vaccine
PG	Peptidoglycan
PMCT	Prevention of mother to child transmission
PMQR	Plasmid mediated quinolone resistance
QECH	Queen Elizabeth Central Hospital
QRDR	Quinolone resistance determining region
RNA	Ribonucleic acid
SBA	Sheep blood agar
SBI	Severe bacterial infection
SC	Sequence cluster
SNP	Single nucleotide polymorphism
tRNA	Transporter rna
UK	United Kingdom
UTI	Urinary tract infection
WGS	Whole genome sequencing
WTSI	Wellcome Trust Sanger Institute

Chapter one: General Introduction

1.1 Background

Antimicrobial resistance (AMR) is a global public health problem, which presents a huge threat to the treatment of all forms of bacterial infections (WHO, 2014; Laxminarayan et al., 2013). A wide range of bacterial pathogens across the globe are increasingly developing resistance to multiple classes of antimicrobial agents rendering the agents concerned ineffective for the treatment of infections. The effects of AMR are substantial. For example, it is estimated that in Europe, 25,000 people die every year from drug resistant bacterial infections (DRI). In the United States, resistant bacteria are reported to cause 2 million illnesses and 23,000 deaths annually (Laxminarayan et al., 2013). A review of AMR by O'Neil et al, estimated a minimum of 700,000 deaths globally to be caused by DRI annually and that by 2050 this number will rise to ten million (O'Neill, 2014). Bacteria that produce extended spectrum β-lactamases (ESBLs) are of major concern. These pathogens are resistant to broad-spectrum cephalosporins in addition to the narrow spectrum β-lactams (i.e. aminopenicillins) and often they also harbour genes that confer resistance to other classes of antimicrobial agents such as fluoroquinolones, aminoglycosides and cotrimoxazole (Paterson, 2006).

SSA is region with an extremely high burden of severe bacterial infection (SBI), which is therefore a leading cause of morbidity and mortality (Reddy et al., 2010).

Furthermore, there is already widespread multidrug resistance to low cost and commonly available antimicrobial agents such as ampicillin, cotrimoxazole, and chloramphenicol, which has necessitated the use of third generation cephalosporins and fluoroquinolones for the management of SBI in the region. However, increasing proportions of bacterial pathogens that are ESBL producing and fluoroquinolone resistant are now being reported from a number of settings across the region (Leopold et al., 2014). Due to a limited range of available antimicrobial agents, ESBL producing bacterial pathogens in this setting result in infections that are more difficult or in some cases, impossible to treat. However, despite the emerging reports, there is very limited understanding on the epidemiology and molecular explanation of ESBL resistance in this region.

1.2 A brief history of antimicrobial agents and antimicrobial resistance in bacteria

The modern era of antimicrobial agents in bacteria begun with the "somewhat by chance" discovery of penicillin by Alexander Fleming in 1928 (Demain and Sanchez, 2009; Ligon, 2004b; Ligon, 2004a). However, there were major technical challenges associated with developing penicillin as a treatment agent. It was not until early 1940's that it began to be used for treatment of serious infection during world war II, after its successful isolation from mould in 1939 by Ernst Chain and Howard Florey (Ligon, 2004c). Between the time that Fleming had discovered penicillin and its successful isolation from mould, in Germany, Gerhard Domagk and others were

working on developing a synthetic antimicrobial agent, sulphonamide, which became the first antimicrobial agent to be introduced into use in 1935 (Ligon, 2004b; Aminov, 2017). Though not as effective as the naturally occurring penicillin, sulphonamides were still useful in treating various bacterial infections, such as pneumonia, meningitis and gonorrhoea (Aminov, 2017).

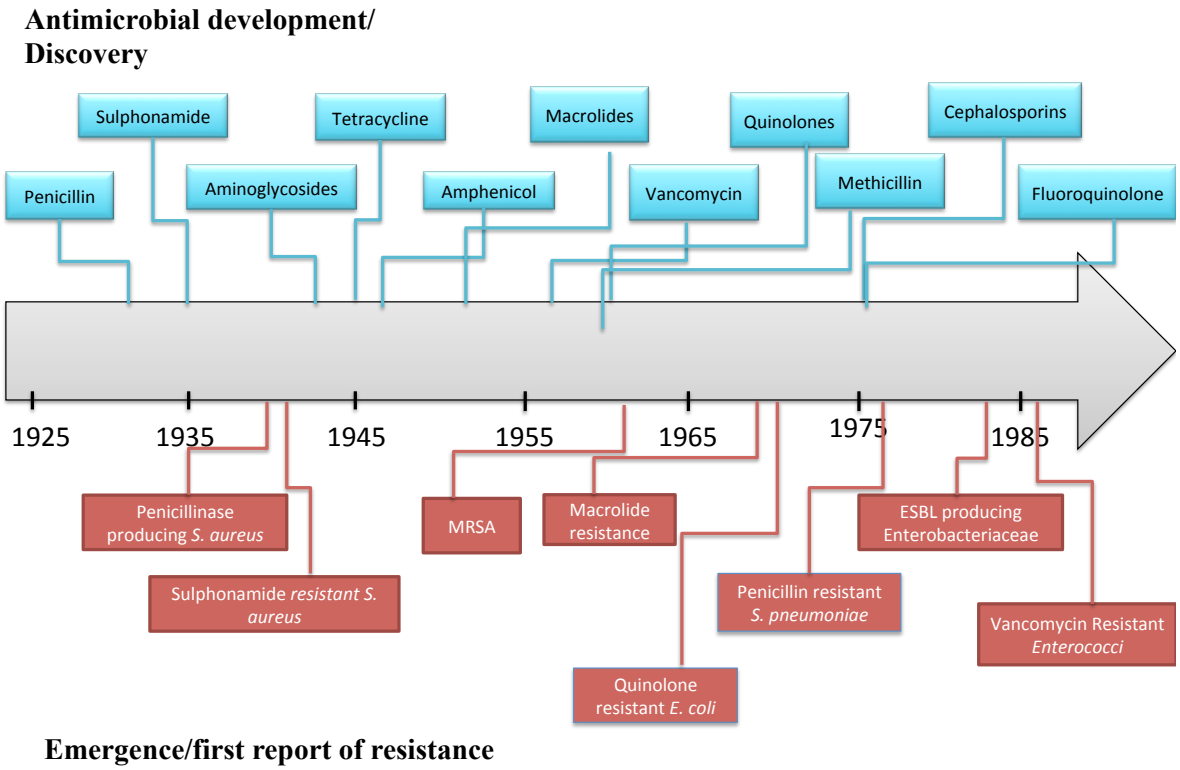

Figure 1.1 Timelines in the discovery/development of antimicrobial agents and emergence of resistance.

In the decades following the introduction of sulphonamides and penicillin, a number of other classes of antimicrobial agents including aminoglycosides (1943),

chloramphenicol (1947), tetracyclines (1945), macrolides (1950) and quinolones (1962) were discovered or developed (Figure 1.1) and introduced into medical use (Powers, 2004; Aminov, 2017). The development of a range of classes of antimicrobial agents offered a variety of options for treating bacterial infections and is considered to be one of the major revolutions of modern medicine that has helped to save millions of lives over the past eight decades.

The introduction of antimicrobial agents into clinical use was however, paralleled by the identification of bacteria that expressed resistance to the agents (Figure 1.1). *Staphylococcus* resistant to both penicillin and sulphonamides were identified as early as the 1940's and by the 1950's penicillin resistance had already become a big problem (Ventola, 2015; Miller and Bohnhoff, 1950). Bacteria expressing resistance to the other classes of antimicrobial agents were subsequently detected. Despite the continued detection of isolates expressing resistance to the early antimicrobial agents, AMR was not perceived to be a major problem due to the speed at which novel classes of antimicrobials were being introduced.

Discovery of novel classes of antimicrobial agents began to stagnate in the late 1960's, but this was still not perceived as a problem due to an increasing ability to make chemical modifications of existing agents (Powers, 2004). The first of these was a β-lactam, methicillin. Introduced in 1960, methicillin was a modification of penicillin that conferred stability to the β-lactamase produced by *S. aureus*. However, the introduction of methicillin was also shortly followed by the identification of methicillin resistant *Staphylococcus aureus* (MRSA) in both the

United Kingdom (1962) and the United States (1968) (Ventola, 2015). This pattern of development of antimicrobials (i.e. vancomycin, cephalosporins and fluoroquinolones) followed by detection of resistance continued through the early 1970s and 1980s (Ventola, 2015). From the late 1980s onwards, the development of new agents has substantially slowed down whilst the spread of AMR pathogens has not (Demain and Sanchez, 2009; Powers, 2004). With some bacterial pathogens becoming resistant to almost all available antimicrobial agents, the world now faces the threat of a return to the pre-antibiotic era.

1.3 Mechanisms of action of antimicrobial agents

In order to understand better how the mechanisms of AMR in bacteria and how bacteria acquire these mechanisms it is useful to also understand the mode of action of antimicrobial agents on bacterial cells. Here, antimicrobial agents have been classified into three groups based on their mode of action through which the agents induce bacterial cell death (bacteriacidal) or growth inhibition (bacteriostatic). These three mechanisms of action include cell wall synthesis inhibition, protein synthesis inhibition and nucleic acid (deoxyribonucleic acid [DNA]) or ribonucleic acid [RNA]) inhibition (Figure 1.2) (Kohanski et al., 2010).

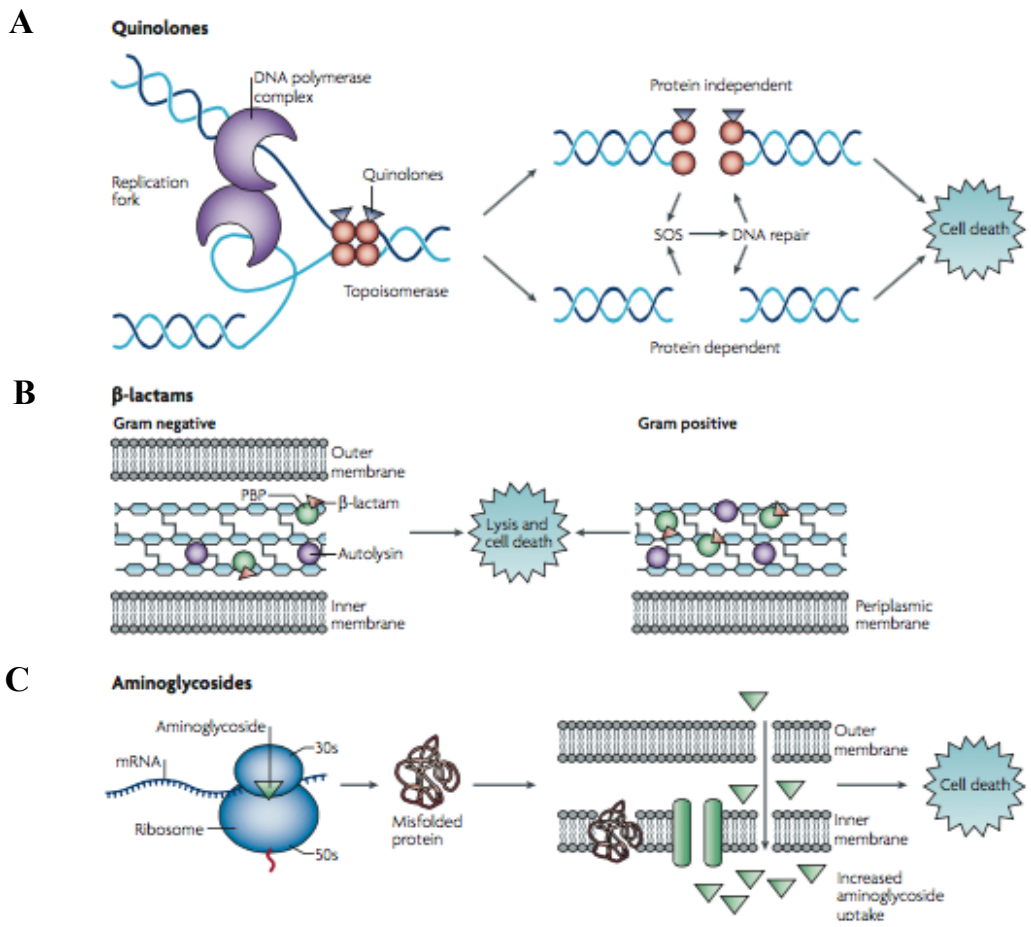

Figure 1.2 An illustration of Antimicrobial (quinolone, β-lactam and aminoglycoside) mechanisms of action that lead to bacterial cell death. (A) Quinolones interfere with changes DNA with the gyrase and topoisomerase IV leading to formation of double-stranded DNA breaks and cell death. (A-II) β-lactams inhibit transpeptidation by binding to PBPs on maturing peptidoglycan strands resulting in decreased peptidoglycan synthesis and increase in autolysins leads to lysis and cell death. (C) Aminoglycosides bind to the 30S subunit of the ribosome and cause misincorporation of amino acids into elongating peptides (Kohanski et al., 2010).

1.3.1 Cell wall synthesis inhibitors

The bacterial cell wall, a polysaccharide structure built from a covalently cross-linked polymer matrix (Figure 1.3) called peptidoglycan (PG), is an important component of most bacterial cells that is responsible for protecting the cell's cytoplasmic membrane from osmotic rapture (Cho et al., 2014). Two types of enzymes provide maintenance to the PG transglycosylase and transpeptidase (also known as penicillin binding proteins, PBPs). Transglycosylase extend the glycan strands of existing PG molecules through addition of disaccharide pentapeptides, whereas PBPs cross-link adjacent peptide strands of immature PG units (Kohanski et al., 2010).

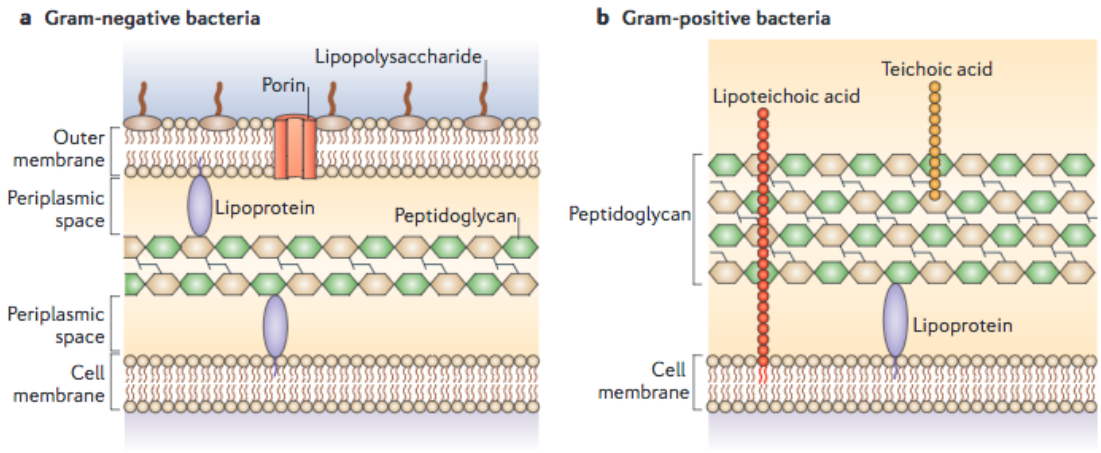

Figure 1.3 Cell walls of Gram-negative (A) and Gram-positive bacteria (Brown et al., 2015)

Inhibition of the cell wall synthesis is the most common mode of action for

antimicrobials on bacterial cells. Common classes of antimicrobial agents that inhibit cell wall synthesis include β-lactams and glycopeptides. Antimicrobial agents belonging to these classes interfere with the bacterial cell wall resulting in changes of cell shape and size or cellular stress responses, all of which lead to cell lysis (Kohanski et al., 2010).

β-Lactams, a class of antimicrobials characterised by possession of a β-lactam ring, which include penicillins, cephalosporins, monobactams, carbapenems, oxapenams, and cephamycins are among the most widely used antimicrobial agents globally (Pitout et al., 2005). Inhibition of bacterial cell wall synthesis by β-Lactams is achieved by covalent bonding of the β-lactam ring to the PBPs, which disrupts the cross-linking of the peptide strands (Figure 1.2B) and hence assembly of the PG (Jovetic et al., 2010). On the other hand glycopeptides such as vancomycin, bind to substrates important for cell wall synthesis and disrupt rigidity to the cell wall (Reynolds, 1989).

1.3.2 Protein synthesis inhibitors

Another common mode of antimicrobial agents is inhibition of protein synthesis. Antimicrobials that inhibit protein synthesis do so by inhibiting either the 50S or 30S ribosomal sub-units, which are important for the translation of messenger RNA (mRNA) to proteins.

50S ribosome inhibitors include such classes of antimicrobial agents as macrolides, (e.g., erythromycin), and amphenicols (e.g.,chloramphenicol) and oxazolidinone

(e.g., linezolid). The 50S ribosome inhibitors work by blocking access of transporter RNA (tRNA) to the ribosome subsequently preventing either initiation of protein translation or translocation of peptidyl-tRNAs (Kohanski et al., 2010).

Common classes of antimicrobial agents that act by inhibiting the 30S ribosomal unit include tetracyclines and aminoglycosides. Tetracyclines bind to the 30S ribosomal subunit, blocking the attachment of the 30S acceptor site with tRNA to prevent protein synthesis (Chopra and Roberts, 2001). Aminoglycosides, which are polycationic hydrophilic sugars with binding affinity to nucleic acids, (Nikaido, 2009) also bind to the 30S ribosome subunit, but specifically the 16S ribosomal RNA (rRNA) component. The bonding of an aminoglycoside and 16S rRNA brings an alteration to the conformation of the complex formed between an mRNA codon and its corresponding charged tRNA at the ribosome (Figure 1.2C). This alteration promotes tRNA mismatching resulting in protein mistranslation (Davies and Davies, 2010; Kohanski et al., 2010).

1.3.3 Nucleic acid synthesis inhibitors

A number of different classes antimicrobial agents including fluoroquinolones, trimethoprim, sulphonamides and rifampicin inhibit the synthesis of either DNA or RNA to induce bacterial cell death.

Quinolones, which have fluoroquinolones as the most important members of the class, act by inhibiting DNA replication whereby they bind to the DNA gyrase and topoisomerase IV enzymes, trapping the enzymes at the DNA cleavage stage (Figure

1.2A) and preventing strand rejoining (Hawkey, 2003).The trapping of gyrase and topoisomerase IV enzymes leads to inhibition of DNA synthesis which subsequently leads to cell death (Kohanski et al., 2010).

Both trimethoprim and sulphonamides work by interfering with the biosynthesis of folic acid, which is essential for thymidine pathway and hence DNA synthesis (Vouloumanou et al., 2011). Sulphonamides target the dihydropteroate synthase enzyme to inhibit synthesis of dihydrofolic acid, whereas trimethoprim binds to dihydrofolate reductase enzyme to prevent the reduction of dihydrofolate to tetrahydrofolate (Alekshun and Levy, 2007; Masters et al., 2003). Cotrimoxazole is a combination of sulphonamide and trimethoprim with a synergistic effect to inhibit the synthesis of folic acid (Smith and Powell, 2000). It is a broad-spectrum antimicrobial agent and one of the most widely used especially in SSA where long-term cotrimoxazole prophylactic therapy is a key component of HIV management programmes.

Rifampicin a key agent in tuberculosis therapy, is a abroad spectrum agent whose mode of action is through non-covalently binding to the active site of a subunit of RNA polymerase to inhibit DNA transcription (Campbell et al., 2001).

1.4 Antimicrobial agents used for treatment of SBI in Malawi

Malawi is a resource-constrained country (see Chapter Two section 2.2.1 for a brief on Malawi's economy) with a very small repertoire of antimicrobial agents available for treatment of bacterial infections. According to the Malawi Ministry of Health

(MoH) Standard Treatment Guidelines (MoH, 2009), in most cases Benzyl penicillin/amoxicillin and chloramphenicol are used as first line antimicrobial agents. Cotrimoxazole is also used as first-line agent but mostly for infections caused by Gram-positive pathogens, as there is widespread cotrimoxazole resistance among Gram-negative pathogens. Ciprofloxacin, ceftriaxone and gentamicin are recommended as reserve (alternative) antimicrobials. However, with reported widespread resistance to the first line agents, the reserve antimicrobials are in many cases used as first line agents (Makoka et al., 2012; Feasey et al., 2015c).

1.5 Multidrug resistance

Until 2011 there was no consensus definition of MDR. However, in 2011 a group of experts in a joint initiative by the European Centre for Disease Control and Prevention (ECDC) and the Centres for Disease Control and Prevention (CDC) defined MDR bacteria as bacteria associated with HA-infection including non-*Salmonellae* Enterobacteriaceae, *Staphylococcus aureus*, *Enterococcus spp.*, *P. aeruginosa* and *Acinetobacter,* resistant to at least one agent in not less than three classes of antimicrobial agents (Magiorakos et al., 2012). The restriction of this definition to bacteria HA infection associated only raises the possibility of inconsistent application in how the term MDR is applied, precluding reliable comparisons of surveillance data including in major pathogens in SSA such as *Salmonellae* and *Streptococcus pneumoniae.* In this thesis, I have adopted the above definition to apply to all pathogens considered.

1.6 Molecular mechanisms of AMR to agents in use in Malawi

1.6.1 Acquisition of AMR

There are several mechanisms through which bacteria express resistance to antimicrobial agents. Some of the major mechanisms include inactivation of the antimicrobial agent, alteration of target site and reduction of the antimicrobial agent concentration in the bacteria cell through restricted cell permeability or efflux pumping. These mechanisms are either intrinsic or acquired (Nikaido, 2009). Intrinsic resistance describes resistance that is naturally attributable to a particular species, because all organisms of that species may lack the appropriate drug-susceptible target or possess barriers that prevent the antimicrobial agent from accessing its target (Kaye et al., 2004). *Pseudomonas aeruginosa* is an example of a bacteria species that is intrinsically resistant to a number of antimicrobial agents such as ceftriaxone, chloramphenicol and tetracycline due to its restricted outer membrane permeability and other mechanisms such as energy-dependent multidrug efflux pumps and chromosomally encoded β-lactamases (Normark and Normark, 2002; Li et al., 1994). Acquired resistance results from a change in the genetic composition of the pathogen so that a drug that once was effective is no longer so. The genetic changes that give rise to genotypes encoding AMR occur through mutations or horizontal gene transfer (Kaye et al., 2004).

1.6.1.1 Horizontal Gene Transfer and Antibiotic resistance

A majority of resistance mechanisms are expressed by genes encoded on extrachromosomal mobile genetic elements (MGEs, [Figure 1.4B]). The most common and largest of such extrachromosomal elements are plasmids. Plasmids are self-replicating double-stranded circular DNA material that can move from one bacterial cell to another in three possible ways including conjugation, transduction or transformation (Figure 1.2A) (Didelot and Maiden, 2010).

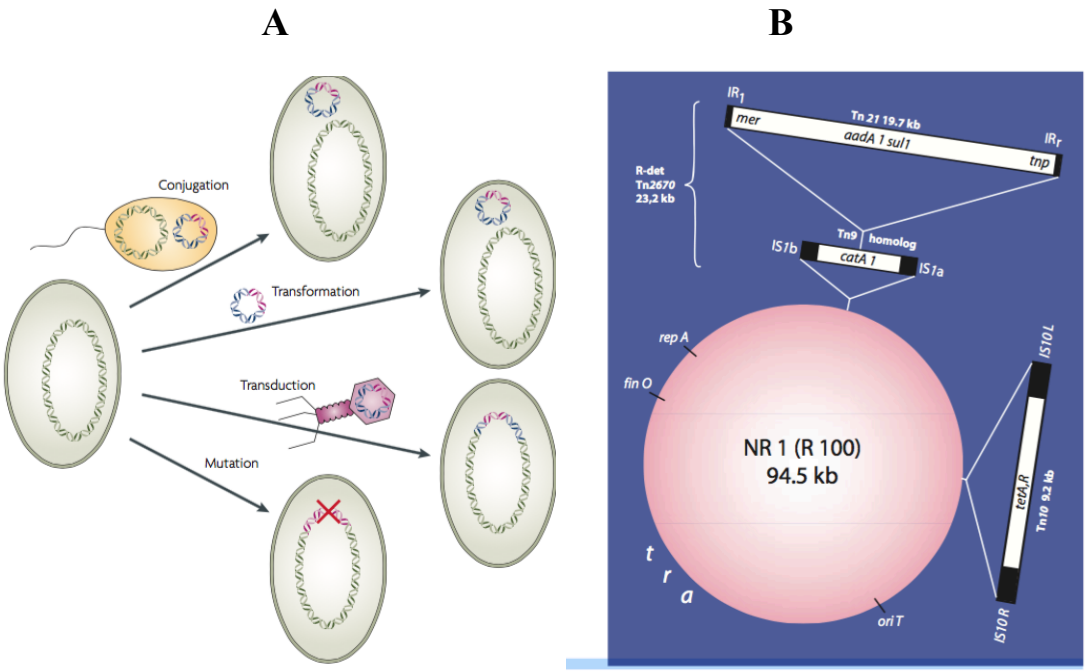

Figure 1.4 An Illustration of mechanisms of acquiring AMR genes and MGEs. (A)Mechanism of DNA. (B) Self-transmissible and multiple resistant R-plasmid with inserts representing mobile transposable elements harbouring some AMR genes (Normark and Normark, 2002).

Plasmid-borne AMR genes are often located within other MGEs carried on the plasmids such as transposons and integrons (Figure 1.4B). Integrons may often be found within transposons, but may also be harboured on plasmids or on chromosomes free from transposons (O'Brien, 2002). Transposons and integrons carry genes that encode enzymes called transposase and integrase respectively and these enzymes enable the corresponding MGE to recombine into chromosomes thereby facilitating the transfer of genetic material including AMR genes between plasmids and chromosomes (Mazel, 2006; O'Brien, 2002). Multiple resistance transposons can also be clustered together to form larger composite transposable elements which can simultaneously transfer multiple unrelated resistance genes (Nikaido, 2009).

1.6.2 Mechanism of β-lactamase resistance

Production of β-lactamases is the major mechanisms through which bacteria express resistance to antimicrobial agents of the β-lactam class. β-Lactamases are enzymes that confer resistance by hydrolysing the β-lactam ring of the agent (Figure 1.5) and hence inactivating its ability to bind to the PBPs (Munita and Arias, 2016; Bonnet, 2004). β-lactamases are grouped into four groups, designated A, B, C and D using the Ambler classification scheme. The Ambler scheme classifies ESBLs based on their amino acid sequence similarity (Bonnet, 2004; Sturenburg, 2003). Classes A, C, and D are enzymes that catalyse hydrolysis of the lactam ring through an acyl-intermediate of serine, whereas the class B enzymes require a zinc metal cofactor

(Hall and Barlow, 2005; Perez-Llarena and Bou, 2009; Shah, 2004). β-Lactamases are more frequently expressed in Gram-negative bacteria. The earliest common β-lactamases including the, Temoniera (TEM-1, TEM-2), and sulfhydryl variable (SHV-1) types belong to the Ambler class A (Sturenburg, 2003). These enzymes confer resistance to aminopenicillins (i.e. amoxicillin) and early, but not later, generation cephalosporins (Paterson, 2006; Nikaido, 2009).

Following the emergence and spread of β-lactamase resistance, extended-spectrum cephalosporins (i.e. third generation cephalosporins) were introduced in the early 1980s to provide alternative treatment options to bacterial infections caused by pathogens with β-lactamase resistance (Paterson, 2006). However, a few years after their introduction, bacterial strains producing ESBLs emerged. ESBLs initially emerged as a result of mutations in the genes encoding narrow spectrum TEM-1, TEM-2, or SHV-1 β-lactamases (Roy et al., 2013; Bonnet, 2004). The mutations modified the narrow-spectrum β-lactamases into new enzymes that were able to hydrolyse cephalosporins, penicillins, monobactams but not cephamycins or carbapenems (Pitout et al., 2005; Paterson, 2006).

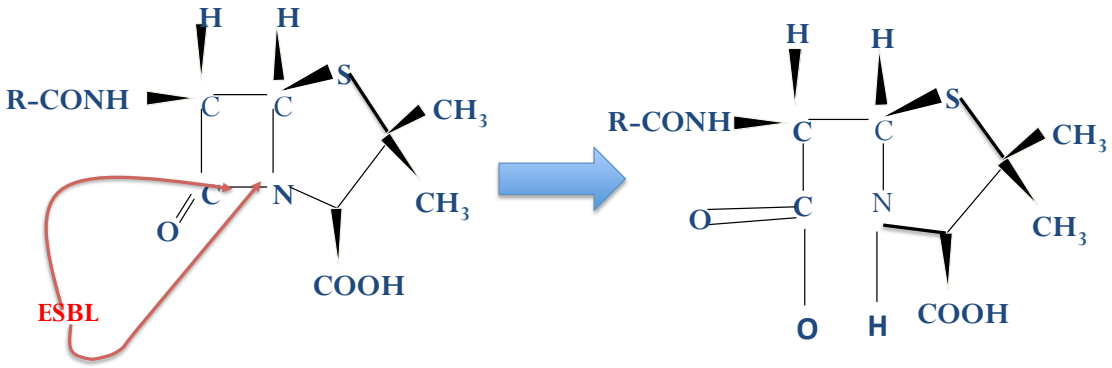

Figure 1.5 An illustration of ESBL hydrolysis of the β-lactam ring. An ESBL enzyme attaches itself to the beta-lactam ring of an antimicrobial agent, breaks the ring bonding and so disables the agent from attaching to the bacterial cell wall.

1.6.2.1 Types of ESBLs

The derivatives of the SHV and TEM β-lactamase types, whose encoding genes are mutants of the *bla*$_{SHV-1}$ and *bla*$_{TEM-1}$ and *bla*$_{TEM-2}$ genes, belong to the Ambler class A (Sturenburg, 2003). These were the earliest known and dominant ESBLs globally in the late 20th century (Bonnet, 2004). The SHV and TEM ESBLs were most commonly detected in isolates of *K. pneumoniae*, and therefore typically associated with hospital acquired (HA) infections (Paterson, 2006).

In the early 1990s however, a new type of ESBLs called CTX-M, originating from *Kluyvera spp*., emerged in pathogenic bacteria (Bonnet, 2004). The CTX-M ESBLs are non-homologous to the SHV or TEM ESBLs, exhibit high-level activity against β-lactams including the cephalosporins and unlike earlier ESBLs, rapidly disseminated

(Bonnet, 2004; Sturenburg, 2003). Of particular interest has been the variant CTX-M-15, which emerged in the early 2000's but has spread to become the most dominant ESBL type across the globe (Munita and Arias, 2016). Whilst SHV and TEM ESBLs were mostly associated with *K. pneumoniae* nosocomial infection, CTX-M ESBLs have been stably acquired by *E. coli* and are frequently implicated in community acquired DRI (Pitout et al., 2005).

In addition to SHV, TEM and CTX-M ESBLs, another common type of ESBLs is the oxacillinase (OXA) type which belongs to the Ambler class D (Shah, 2004; Sturenburg, 2003). Unlike the SHV, TEM and CTX-M ESBLs, which are frequently found in Enterobacteriaceae, the OXA ESBLs are mostly found in *Pseudomonas* (Paterson, 2006).

The genes encoding ESBLs are frequently, but not exclusively found on plasmids as they can also be integrated onto the chromosome of some isolates. Whether found on plasmid or chromosome, the ESBL genes are usually incorporated into MGEs including transposons or integrons as previously described. These MGEs especially those containing the CTX-M ESBLs, often harbour other genes encoding resistance to multiple antimicrobial agents such as aminoglycosides and sulphonamides. As a result, bacteria with ESBL resistance are usually multidrug resistant (MDR).

1.6.3 Mechanisms of fluoroquinolone resistance

Resistance to fluoroquinolones in bacteria occurs mostly as a result of mutations in the gyrase (*gyrA* and *gyrB*) or topoisomerase IV (*parC* and *parE*) genes (Munita and

Arias, 2016; Hooper and Jacoby, 2015a). The DNA gyrase and topoisomerase IV are essential for bacterial replication through DNA synthesis and fluoroquinolones act by inhibiting these enzymes (Blondeau, 2004). Although both gyrase and topoisomerase mutations can occur in both Gram-negative and Gram-positive bacteria, mutations in the gyrase (*gyrA o*r *gyrB*) genes are mostly associated with fluoroquinolone resistance in Gram-negative bacteria whereas mutations in the topoisomerase (*parC* and *parE*) are mostly responsible for fluoroquinolone resistance in Gram-positive bacteria (Hawkey, 2003). Specific loci of the gyrase or topoisomerase amino acid sequences, called quinolone resistance determinant regions (QRDRs) have been identified as being particularly important in conferring fluoroquinolone resistance (Alekshun and Levy, 2007).

Other chromosomal mutations may also confer resistance to fluoroquinolones by reducing the drug concentration through efflux activity, which pumps the drug out of the bacteria, or by limiting permeability of the drugs into the bacteria cell as a result of porin channel mutations (Nikaido, 2009). In addition to the chromosomal mutations, recent studies have also identified plasmid-mediated quinolone resistance (PMQR) genes that confer low-level resistance to fluoroquinolones. These genes include *qnrA*, *qnrB*, *qnrC, qnrD qnrS, qnrVC* and *aac(6")-lb-cr* genes (Jacoby et al., 2014). The specific mechanisms through which these plasmid-mediated genes confer resistance to fluoroquinolones are not clear but it is thought that *qnr* genes limit the binding of fluoroquinolones to the gyrase genes. Furthermore, *qnr* genes have been understood to be a precursor of *gyrA* mutations, which in turn lead to high-level fluoroquinolone resistance (Jacoby et al., 2014).

1.6.4 Mechanism of resistance to aminoglycosides

Resistance to aminoglycosides is through inactivation of the antimicrobial agent by bacterial enzymes that reduce the net positive charge of the aminoglycoside, hence affecting its affinity to the bacterial RNA (Nikaido, 2009; Mingeot-Leclercq et al., 1999). A number of different enzymes confer aminoglycoside resistance and these include aminoglycoside acetyltransferase (AAC), phosphoryltransferase (APH), and aminoglycoside adenyltransferase or nucleotidyltransferase (ANT). Each of these classes of enzymes also has several variants (Nikaido, 2009). Four acetyltransferases [AAC(2'), AAC(6'), AAC(1), and AAC(3)], seven phosphotransferases [APH(3'), APH(2"), APH(3"), APH(6), APH(9), APH(4), and APH(7")], four nucleotidyl-transferases [ANT(6), ANT(4'), ANT(3"), and ANT(2")], and a bifunctional enzyme, AAC(6')–APH(2"), conferring aminoglycoside resistance have been identified previously (Kotra et al., 2000).

Whilst the enzymatic inactivation of aminoglycosides is considered the main mechanism of resistance, bacterial resistance to aminoglycosides can also be expressed through reduction of drug concentration in the bacterial cell through efflux activity (Mingeot-Leclercq et al., 1999).

1.6.5 Mechanism of resistance to chloramphenicol

The common mechanism of chloramphenicol resistance is drug inactivation by acetyltransferase enzymes (Alekshun and Levy, 2007; Schwarz et al., 2004). The most common acetyltransferase enzymes conferring resistance to chloramphenicol

are in two non-homologous proteins designated chloramphenicol acetyltransferase A (Type A CAT) and chloramphenicol acetyltransferase B (Type B CAT). Although the *cat* genes encoding type A CAT (*catA*) or type B CAT (*catB*) can be found on a chromosome sequence, they are mostly plasmid borne and are usually incorporated on a transposon element (Schwarz et al., 2004).

1.6.6 Mechanism of resistance to cotrimoxazole

Resistance to sulphonamides and trimethoprim is through mutations in the dihydropteroate synthase and dihydrofolate reductase genes resulting in reduced affinity of sulphonamides and trimethoprim to the dihydropteroate synthase and dihydrofolate reductase genes respectively (Alekshun and Levy, 2007).

1.7 AMR in major bacterial pathogens in sub-Saharan Africa

1.7.1 Severe bacterial infections in SSA

When considering antimicrobial resistance, it is also important to first consider the very real need for these agents in SSA; more people will die from lack of access to appropriate antimicrobial chemotherapy in the next few years than from DRI (Laxminarayan et al., 2016). Accordingly, the causes and consequences of severe bacterial infections (SBI) including bloodstream infection (BSI), lower respiratory infection and meningitis in SSA are briefly reviewed here.

Whilst bacterial pathogens are able to cause infections in healthy individuals, in SSA, there is a high prevalence of immunosuppressive conditions which favour the

development of severe and invasive bacterial disease, including HIV/AIDS, malaria and malnutrition (Feasey et al., 2015a; Scott et al., 2011; Holmes et al., 2003). According to the UNAIDS Global AIDS Update for 2016, 25.5 million people were estimated to be HIV infected in SSA representing 69.8% of the global HIV prevalence in the year 2015 (UNAIDS, 2016) SSA is also a malaria endemic region. Two hundred fourteen million cases of malaria were estimated to have occurred globally in 2015 and 90.0% of these cases were in Africa (mostly SSA) (WHO, 2016)The majority of severe malaria cases occur in children aged below five years whilst HIV is highly prevalent among young to middle-aged adults. Malnutrition is also common among children under five years old. These factors have in turn created huge sub-populations of persons that are more vulnerable to SBI.

SBI are among the leading causes of mortality and morbidity in Africa (Seale et al., 2014; Reddy et al., 2010). BSI is the most common manifestation of SBI in SSA with non-typhoidal *Salmonella* (NTS) and *Streptococcus pneumoniae* identified as the leading causes of BSI in SSA in the past two decades, but recently there has also been an upsurge *S.* Typhi BSI (Feasey et al., 2015c; Al-Emran et al., 2016; Reddy et al., 2010). Overall mortality rate of BSI in SSA was previously estimated at 18.1% (Reddy et al., 2010) but among children aged below 5 years, BSI mortality rates of up to 27% have previously been estimated (Gordon et al., 2008; Morpeth et al., 2009). In HIV infected adults, BSI mortality rates as high as 47% have also been reported (Huson et al., 2014; Morpeth et al., 2009).

Another common manifestation of SBI is pneumonia. Pneumonia ranks amongst the top two leading causes of child death, which also includes preterm birth (Liu et al., 2012). In 2015 for example, pneumonia was responsible for 15% of all child deaths globally (Unicef, 2017). As with other infectious disease syndromes, the majority of the pneumonia cases and deaths occur in developing countries with SSA having the highest incidence of child pneumonia alongside South East Asia (Walker et al., 2013). Pneumonia is also a common cause of death in adults, especially those that are HIV infected. In Ethiopia for example, pneumonia was identified as one of the common causes of death in both adults and children (Abejew et al., 2014). *S pneumoniae* is the leading bacterial cause of pneumonia, and other bacterial causes include *Haemophilus influenza* type b, *Staphylococcus aureus and Klebsiella pneumoniae* (Rudan et al., 2008; Walker et al., 2013).

Meningitis is another bacterial disease syndrome to which patients are more susceptible when HIV infected or in early childhood (Molyneux et al., 2009; Gordon et al., 2000). The most common causes of bacterial meningitis in SSA include *S. pneumoniae*, *H. influenza, Neisseria meningitis* (the meningococcus) (Veltman et al., 2014). NTS has also been identified as common cause bacterial meningitis in some settings and similar to NTS BSI, NTS meningitis mostly affects HIV/AIDS infected individuals and children, but children suffer the highest burden (Molyneux et al., 2009; Wall et al., 2014). Bacterial meningitis has very high mortality rates in both children and adults. *S. pneumoniae*, *H. influenzae* and the meningicoccus meningitis were previously associated with overall mortality rate of 27.4% in African children (Peltola, 2001) whereas amongst adults in Malawi and Senegal, pneumococcal

meningitis has been associated with mortality rates above 50.0% (Veltman et al., 2014; Gordon et al., 2000; Wall et al., 2013).

1.7.2 AMR in *Salmonellae*

Both typhoidal and nontyphoidal *Salmonellae* are major causes of BSI in SSA. Invasive NTS (iNTS) disease in SSA is largely, but not exclusively associated with two serovars, *S*. Typhimurium and *S*. Enteritidis. *S*. Typhimurium has however, been the predominant NTS serotype in most SSA settings (Feasey et al., 2015c; Gordon et al., 2008; Abraham and Chain, 1988). In Kenya, *S*. Typhimurium was the predominant isolate in children with *Salmonella* bacteraemia (Kariuki et al., 2006). In South Africa of 1,318 microbiologically confirmed invasive NTS cases between 2003 and 2004, 67.0% were *S.* Typhimurium and 10.0% *S.* Enteritidis while in Malawi, of 10,139 invasive NTS cases isolated between 1998 and 2014, 8,017 (79.1%) were *S.* Typhimurium and 1,608 (15.8%) *S.* Enteritidis (Feasey et al., 2015c). Both *S*. Typhimurium and *S*. Enteritidis are associated with distinct African variants. For example, *S.* Typhimurium of multi locus sequence type (MLST) 313 has been implicated in driving a number of NTS epidemics in SSA (Okoro et al., 2012). The picture regarding *S.* Typhi is less straight forward; whilst global burden estimates have suggested it to be a common cause of BSI in SSA, it is only after 2010 that there have been a number of reports of epidemics of Typhoid fever in Southern and Eastern Africa. The global burden estimates have been based on few, and sometimes historic studies including three from Egypt and South Africa (Mogasale et al 2014) hence it is indeed likely that, prior to 2010, prevalence of *S*. Typhi in SSA

was genuinely low (Feasey et al., 2015c; Mandomando et al., 2010). Recent studies however, have shown huge increases in typhoid infections across SSA, with some reported epidemics especially after 2010 (Feasey et al., 2015b; Nelson, 2012). Evidence from studies that had access to WGS revealed the expansion of the H58 clade, which is particularity associated with AMR (Wong et al., 2015; Hendriksen et al., 2015).

The majority of both NTS and *S*. Typhi isolates causing severe infection in SSA are MDR (resistant to ampicillin, cotrimoxazole, and chloramphenicol) (Mandomando et al., 2010; Feasey et al., 2015c; Leopold et al., 2014). In the review by Reddy et al, NTS isolates (both *S*. Typhimurium and *S*. Enteritidis) were commonly resistant to ampicillin, chloramphenicol and cotrimoxazole, although there were substantial differences in the susceptibility rates for *S.* Enteritidis and *S.* Typhimurium. For instance, while 100%, 15% and 3% of *S.* Typhimurium isolates were susceptible to chloramphenicol, ampicillin and cotrimoxazole, 38.9%, 33.6% and 93.2% of the *S.* Enteritidis were susceptible to chloramphenicol, ampicillin and cotrimoxazole, respectively. On the other hand, the majority (≥ 89%) of the *S*. Typhi isolates were fully susceptible to these agents (Reddy et al., 2010). Epidemics of both NTS and *S.* Typhi were associated with MDR in Malawi where at one point 93% of *S.* Typhimurium isolates and 97% of *S.* Typhi isolates in 2014 were MDR (Feasey et al., 2015c). A study from an urban setting in Ghana showed 64% resistance of all *Salmonella* isolates to ampicillin, 52.7% to cotrimoxazole, 73.9% to chloramphenicol and 67% to tetracycline with over 40% of all isolates of significant pathogens identified in this study being MDR (Obeng-Nkrumah et al., 2016). A majority of NTS

isolates from children (< 15 years of age) in Mozambique were resistant to ampicillin (75.0%), chloramphenicol (55.0%) and cotrimoxazole (66.0%) (Mandomando et al., 2010).

The incidence of MDR invasive *Salmonella* disease in SSA has been a key driver to use of cephalosporins and fluoroquinolones as antimicrobials of choice for the empirical treatment of suspected bacterial infections (Feasey et al., 2015c; Kariuki et al., 2010). Although reports of cephalosporin and fluoroquinolone resistance have begun to increasingly emerge across the region, high proportions of both NTS and *S*. Typhi remain susceptible to third generation cephalosporins and fluoroquinolones with most countries reporting isolated or no cases of cephalosporin or fluoroquinolone resistant isolates (Leopold et al., 2014; Feasey et al., 2015c; Mandomando et al., 2010). In one study from a rural setting in Ghana, no ESBL resistance was detected in NTS (n=215) or *S*. Typhi (n=110), whereas resistance to ciprofloxacin was detected in 6.5% NTS and 0.0% of *S.* Typhi (Eibach et al., 2016). Another study from urban Ghana also reported 100% and 97.6% susceptibility to ciprofloxacin and levofloxacin among *Salmonella* isolates but substantial proportions of isolates with resistance to third generation cephalosporins (ceftriaxone [16.7%] and cefotaxime [4.8%]) (Obeng-Nkrumah et al., 2016). Only 19 isolates (from 10,139 NTS and 2054 *S.* Typhi isolates) were reported to have altered susceptibility to ciprofloxacin in Malawi of which one *S*. Typhimurium isolate was confirmed to be fully ciprofloxacin resistant and one *S.* Typhi isolate was confirmed to have diminished ciprofloxacin susceptibility (DCS). The same study also identified 16 *S*. Typhimurium isolates with resistance to third generation cephalosporins

(Feasey et al., 2015c). In Mozambique, all 307 and 48 isolates were tested for susceptibility to fluoroquinolones and cephalosporins respectively, were susceptible. Reduced susceptibility to fluoroquinolone in *Salmonella* isolates has also been reported in a number of other countries such as the Democratic Republic of Congo, Zambia and Kenya (Lunguya et al., 2012; Smith et al., 2010; Al-Emran et al., 2016; Hendriksen et al., 2015).

In summary, whilst there have been isolated cases of resistance to cephalosporins and fluoroquinolones, *Salmonellae* in Africa are mostly resistant to the first-line agents. Because *Salmonella* is a very common cause of bacteraemia in SSA, iNTS infections continue to drive the empirical use of ceftriaxone and fluoroquinolones.

1.7.3 AMR in *Escherichia coli*

Escherichia coli is a member of the human and mammalian gastrointestinal tract microbiota, containing both commensal strains as well as pathogenic strains that can cause potentially life-threatening infections (Kaper et al., 2004; Johnson and Russo, 2002). Pathogenic *E. coli* are categorised as either intestinal *E. coli* (or diarrheagenic *E. coli* [DEC]) or Extra-intestinal pathogenic *E. coli* (ExPEC) depending on the disease syndrome with which they are associated (Jonson, 2005; Russo and Johnson, 2003; Kohler and Dobrindt, 2011).

In Africa, *E. coli* is the second most common cause of community acquired (CA) Gram-negative BSI accounting for 7.3% of all confirmed CA-BSI isolates (Reddy et

al., 2010). Alongside *S. aureus*, *K. pneumoniae* and *P. aeruginosa*, *E. coli* has also been identified in a number of countries within the region as a cause of HA infection through a wide range of disease syndromes such as urinary tract infection (UTI), BSI, surgical wound infection and acute respiratory infection (Leopold et al., 2014; Reddy et al., 2010; Maina et al., 2016).

Life threatening invasive *E. coli* disease is particularly linked to *E. coli* that are becoming increasingly resistant to available antimicrobial agents, especially those of ST131 (Pitout, 2012). Globally, *E. coli* is one of the pathogens with high rates of resistance to a wide range of antimicrobial agents (WHO, 2014). Antimicrobial susceptibility data from across SSA indicate that very high proportions of *E. coli* isolates from the region are MDR and antimicrobial agents such as ampicillin and cotrimoxazole are rarely effective in treating *E. coli* infections while high rates of resistance to chloramphenicol are common. Reddy et al (Reddy et al., 2010) reported non-susceptibility rates of 91.1%, 83.1% and 76.6% against ampicillin, cotrimoxazole and chloramphenicol respectively detected in *E. coli* isolates from across Africa while gentamicin resistance was 19.8%. Of more concern however, is that clinical isolates of *E. coli* with resistance to cephalosporins and fluoroquinolones have become increasingly common in the last decade (Leopold et al., 2014). Reddy *et al* found 100% susceptibility to ceftriaxone and ciprofloxacin in *E. coli* isolates but other contemporaneous studies from SSA revealed that cephalosporin and fluoroquinolone resistant *E. coli* was being detected in a number of African countries. For example, one study in Tanzania (Blomberg et al., 2005) detected cephalosporin resistance in 25.0% *E. coli* isolated from paediatric blood

culture (2001 and 2004). Fluoroquinolone and cephalosporin resistant *E. coli* isolates were also detected in Cameroon from *E. coli* isolated between 1995 and 1998 (Gangoue-Pieboji et al., 2005a; Gangoue-Pieboji et al., 2006) and in Malawi, from cultures isolated in 2004-2005. Recent data show that cephalosporin and fluoroquinolone resistance has increased in the last decade. A systematic review of AMR in Africa by Leopold *et al.* (Leopold et al., 2014) show that the overall median proportion of cephalosporin resistance in *E. coli* isolates from patients with febrile illness in SSA in 2012 ranged between 0.0-80.0% although the study which reported 80% had only 10 isolates. Excluding this study, the median range of proportions of isolates with cephalosporin resistance was 0.0% to 37.0%, fluoroquinolone resistance from 0% to14.3% and gentamicin resistance from 0.0% to 35.0%. More recent studies reveal higher proportions of *E. coli* isolates that are resistant to third generation cephalosporins, fluoroquinolones and aminoglycosides. A retrospective analysis of blood culture isolates from January 2010 to December 2013 in Ghana revealed increased resistance to cefuroxime (70.2%), cefotaxime (56.2%), ciprofloxacin (66.1%), levofloxacin (71.0%) and gentamicin (54.0%) (Obeng-Nkrumah et al., 2016). Cephalosporin and fluoroquinolone resistant *E. coli* have also been reported from other countries in SSA such as Angola, Nigeria, South Africa and Tanzania (Tansarli et al., 2013; Mshana et al., 2016; Ribeiro et al., 2016; Dramowski et al., 2015; Peirano et al., 2011; Raji et al., 2015).

1.7.4 AMR in *Klebsiella pneumoniae*

K. pneumoniae is the most medically important species of *Klebsiella*, a genus that is ubiquitously found in the environment and mucosal surfaces of mammals (Podschun and Ullmann, 1998). In humans, *K. pneumoniae* can be carried asymptomatically in the intestinal tract, skin nose and throat of health individuals, but it can also be an opportunistic pathogen and able to cause a variety of both HA infection and CA infections such as sepsis, pneumonia, wound, soft tissue or UTI, especially among individuals with compromised immune system (Podschun and Ullmann, 1998; Jones, 2010). Globally, *K. pneumoniae* is recognised as one of the leading causes of HA infection, which are associated with high mortality rates even where effective antimicrobial therapy is available (Podschun and Ullmann, 1998). Additionally, *K. pneumoniae* are recognised for having a particular ability to acquire extensive AMR, including to cephalosporins, fluoroquinolones and carbapenems, which are agents of last resort in many settings.

In SSA, there is a paucity of data on infections caused by *Klebsiella spp*. but available studies show that *K. pneumoniae* cause both HA and CA infections in the region. A study in Nigeria identified *K. pneumoniae* among three most common causes of BSI in infants alongside *E. coli* and *S. aureus* and was reported to be responsible for 10.3% of the infant BSI cases (Ayoola et al., 2003; Ayoola et al., 2005). Similarly in Kenya, Maina *et al* (Maina et al., 2016) identified *K. pneumoniae* as the third most common cause of CA-BSI (8.0% of all cases). In the same Kenyan study, *K. pneumoniae* was also the second-most common cause (after *Candida spp*.) and the

leading bacterial cause of HA-BSI (Maina et al., 2016). In South Africa, 34.0% of *K. pneumoniae* isolates were associated with CA-BSI and the other 66.0% of *K. pneumoniae* with HA infection. Pneumonia was the most common disease syndrome (62.0%) associated with *K. pneumoniae* in South Africa followed by UTI (10.0%) and meningitis (10.0%), intravascular catheter-related infections (5.0%) and intra-abdominal abscess (5.0%).

Despite limited available data on *K. pneumoniae* with ESBL and fluoroquinolone resistance in SSA, recent studies have shown that *K. pneumoniae* strains with ESBL and fluoroquinolone resistance have emerged and are spreading rapidly in the region (Desta et al., 2016; Ouedraogo et al., 2016). *K. pneumoniae* isolates dominated ESBL Enterobacteriaceae isolates previously identified in Malawi (Gray, 2006). Predominance of *K. pneumoniae* among bacterial isolates expressing ESBL resistance has also been reported in a number of other countries in SSA such as Tanzania, Kenya, Ethiopia, Burkina Faso and Ghana among others within the region (Eibach et al., 2016; Ouedraogo et al., 2016; Mshana et al., 2013a; Desta et al., 2016). In Ghana, *K. pneumoniae* accounted for 82.9% (34/41) of ESBL producing Enterobacteriaceae and similarly in Ethiopia, *K. pneumoniae* was the leading ESBL producer with 75.9% of *K. pneumonia* isolates expressing ESBL resistance phenotype.

Fluoroquinolone resistance is also more common in *K. pneumoniae* than other bacterial pathogens. However, there are substantial variations in the proportions of *K. pneumoniae* isolates expressing resistance to fluoroquinolones among different

settings and time points. Leopold et al (Leopold et al., 2014) reported low fluoroquinolone resistance rates in the range 0.0%-4.8% among *K. pneumoniae* isolates for studies published between 1994 and 2007 (Obi et al., 1996; Adeyemo et al., 1994; Moges et al., 2002; Tessema et al., 2007). In Malawi only 2/649 Enterobacteriaceae isolates showed resistance to ciprofloxacin and both isolates were *K. pneumoniae* (Gray, 2006). However, higher rates of resistance have been reported in other settings especially in recently published studies. *K. pneumoniae* had the highest prevalence of ciprofloxacin resistance in Ghana (39.0%) among ESBL producing Enterobacteriaceae causing BSI isolated between 2007 and 2012 (Eibach et al., 2016). Another recent report showed similar proportions of *K. pneumoniae* isolates (39.9% and 36.8%) expressing resistance to ciprofloxacin and levofloxacin respectively), and these proportions were higher than for any other major pathogen (Obeng-Nkrumah et al., 2016). Similar resistance rates were also reported from Kenya from blood isolates (32.0%-48.0%), but here lower proportions (16.0%-24.0%) of isolates from urine samples expressed resistance to ciprofloxacin (Maina et al., 2016). A previous review by Mshana et al had shown higher prevalence of resistance to ciprofloxacin in *K. pneumoniae* isolates from urine samples than isolates from blood samples in Tanzania and Mozambique (Mshana et al., 2013b).

Just as with *E. coli,* other antimicrobial agents commonly available in SSA have become almost ineffective for treating *K. pneumoniae* infections. *K. pneumoniae* is intrinsically resistant to aminopenicillins (Bouza and Cercenado, 2002). Rates of resistance to antimicrobial agents such as cotrimoxazole and gentamicin are as high

as 95% and 67% respectively in *K. pneumoniae* isolates from some countries in the region (Mshana et al., 2013b). The higher levels of resistance in these agents means cephalosporins and fluoroquinolones are the agents of choice for treating *K. pneumoniae* infections across SSA. However, the rapidly increasing ESBL and fluoroquinolone resistance highlighted here also means *K. pneumoniae* infections are becoming untreatable in a number of countries in SSA.

In settings where there has been widespread ESBL and fluoroquinolone resistance, carbapenems, where available, have been the antimicrobials of last resort for infections caused by ESBL producing *K. pneumoniae* (Pitout et al., 2015). This also includes a few countries in SSA. However, carbapenemase producing *K. pneumoniae* which are resistant to carbapenems are now being detected in a number of countries globally (Chang-Ro Lee, 2016). Although use of carbapenems remains uncommon in a majority of countries in SSA, some of the few countries which have been using carbapenems have now begun reporting the emergence and spread of KPC (Poirel et al., 2011; Brink et al., 2012b). The emergence of carbapenem resistance in some countries in SSA raises the possibility that in the absence of appropriate antimicrobial stewardship, carbapenems will have a short period of efficacy before resistance spreads across the region.

1.7.5 AMR in *Streptococcus pneumoniae*

As described in this chapter, *S. pneumoniae* is one of the leading bacterial pathogens in SSA associated with all forms of bacterial disease syndromes including BSI, lower

respiratory infection and meningitis. While there have been reports of up to 30% MDR in some settings in SSA such as Malawi, *S. pneumoniae* remain largely (≥ 90) susceptible to key first line antimicrobial agents particularly penicillin (Everett et al., 2011).

Antimicrobial agents to which *S. pneumoniae* show relatively higher rates of resistance to in SSA include chloramphenicol, cotrimoxazole, tetracycline and erythromycin (Reddy et al., 2010). In Malawi over 90.0% of *S. pneumoniae* were resistant to cotrimoxazole, between 50.0-63.0% to tetracycline and between 20.0-30.0% to chloramphenicol (Everett et al., 2011; Makoka et al., 2012). Two studies from Ghana and South Africa both describing trends in paediatric BSI, reported penicillin non-susceptibility rates of more than 30.0% (Obeng-Nkrumah et al., 2016; Dramowski et al., 2015), but several published studies show that in other parts of SSA, over 90.0% of *S. pneumoniae* isolates remain susceptible to penicillin (Reddy et al., 2010; Everett et al., 2011). *S. pneumoniae* isolates also remain largely susceptible to cephalosporins and fluoroquinolones (Cornick and Bentley, 2012; Everett et al., 2011).

1.7.6 AMR in *Staphylococcus aureus*

S. aureus has been identified as a common cause of infection in both adult and paediatric patients in a number of studies from the region (Makoka et al., 2012; Obeng-Nkrumah et al., 2016; Mandomando et al., 2010; Dramowski et al., 2015). One study on bacterial infections in Malawi with isolates from a wide range of

clinical sources including blood, CSF, joint fluid, bone biopsy, and pus identified *S. aureus* as the most common isolated pathogen from paediatric blood culture, third most common isolate from adult blood culture and overall most common isolated pathogen from abscesses, bones and joints (Makoka et al., 2012). However, two other studies from the same country identified very small proportions (< 5.0%) of *S. aureus* among the bacterial causes of BSI in adults (Feasey et al., 2014b; Gordon et al., 2001). *S. aureus* was also identified as the most common cause of BSI in paediatric patients in Ghana (Obeng-Nkrumah et al., 2016) but second and third most common cause of BSI in pediatric patients in South Africa and Mozambique, respectively (Dramowski et al., 2015; Mandomando et al., 2010).

The majority of isolates of *S. aureus* reported from SSA are resistant to penicillin, cotrimoxazole and chloramphenicol (Mandomando et al., 2010; Obeng-Nkrumah et al., 2016). As previously described, MRSA are *S. aureus* resistant to methicillin, but more importantly they are MDR and mostly associated with HA-infections. Data on MRSA in SSA are scarce, perhaps because the right clinical conditions do not exist to facilitate its spread in hospitals (for example short patient stays, lack of intensive care facilities and lack of long term vascular catheters) in addition to the general lack of surveillance across the region. Published data however, show that the prevalence of MRSA ranges from 7.0% to 36.0% across SSA. In Mozambique, 8.0% of *S. aureus* isolates were methicillin resistant (Mandomando et al., 2010), and in Zimbabwe the proportion was 7.0% (Mauchaza et al., 2016). In South Africa however, as many as 23.0% and 36.0% of *S. aureus* were identified as MRSA (Shittu and Lin, 2006; Fortuin-de Smidt et al., 2015). Over 90.0% of the MRSA isolates from

Zimbabwe and South Africa were also resistant to antimicrobial agents such as gentamicin, cotrimoxazole and erythromycin (Mauchaza et al., 2016; Shittu and Lin, 2006).

1.8 Molecular epidemiology of ESBL and fluoroquinolone resistance in SSA

1.8.1 Molecular determinants of ESBL resistance

Just as there are limited phenotypic antimicrobial resistance data, there are also limited molecular data describing the determinants of AMR in SSA. ESBL producing bacterial were first reported in SSA in the late 1980's when broad spectrum cephalosporin resistant Enterobacteriaceae were first detected in South Africa (Pitout et al., 1998). This was very few years after ESBLs were first reported in Europe between 1983 and 1984 (Pitout et al., 1998; Philippon et al., 1989). The available data further show that TEM and SHV ESBL types were the earliest and dominant ESBLs in SSA until the mid 2000s, just as in many parts of the world. In South Africa, a study investigating mechanisms of resistance to broad-spectrum cephalosporins in *K. pneumoniae*, *E. coli* and *Proteus mirabilis* identified TEM-26, as the dominant ESBL, followed by SHV-2 and SHV-5 (Pitout et al., 1998). In Cameroon, SHV-12 and SHV-2 were the dominant ESBLs in late 1990's (Gangoue-Pieboji et al., 2005b; Gangoue-Pieboji et al., 2005a) and although dominance of SHV and TEM ESBLs continued to be reported, new ESBLs of the CTX-M emerged in the early 2000s. CTX-M-15 was first detected in two *E. coli* and one *K. pneumoniae* from 17

ESBL producing Enterobacteriaceae whereas the other isolates produced SHV-12 (Gangoue-Pieboji et al., 2005b). In Malawi, a polymerase chain reaction (PCR)-based study characterising ESBLs from Enterobacteriaceae, identified the TEM, SHV and CTX-M ESBL types from ten isolates and the SHV ESBLs were the dominant type (SHV-12 [three isolates], SHV-27 [one isolate] followed by TEM-63 (two isolates), where as CTX-M-15 was only detected in one *K. pneumoniae* isolate (Gray, 2006). A higher proportion of TEM-63 (36.8%) ESBLs was reported in Tanzania in 2005 but here, prevalence of CTX-M-15 (31.6%) amongst other ESBLs was higher than the one reported in Malawi (Blomberg et al., 2005). The emergence of CTX-M-15 in the early 2000s was also consistent with its emergence in other parts of the world such as Europe, South America and Asia in the late 1990's and early 2000s (Bonnet, 2004).

Recent studies from SSA now suggest CTX-M-15 is currently the dominant ESBL type with *E. coli* as its most common producer. For example, the first characterisation of ESBLs in Cameroon reported dominance of SHV-12 among the detected ESBLs and a higher prevalence of ESBL *Klebsiella spp.* (18.8%) than *E. coli* (14.3%) (Gangoue-Pieboji et al., 2005a). However, a more recent study describing the molecular epidemiology of ESBL resistance in the same country found CTX-M-15 as the most common ESBL with a prevalence of 96.0% where as SHV-12 was only present in 4.0% in all the ESBL producing Enterobacteriaceae. Although there was a variation in the prevalence of detected ESBL producing organisms between study sites, with *Klebsiella* being more common in one hospital and *E. coli in* two hospitals, overall, there was a higher proportion of ESBL producing *E. coli* than ESBL

producing *Klebsiella* spp.(Lonchel et al., 2012). A meta-analysis of proportion estimates of ESBL producing bacteria in East Africa found six studies which had reported prevalence of ESBL types identified and four of the studies (two from Kenya, one from Tanzania and one from Uganda) had identified CTX-M as the dominant ESBL type while one study (from Tanzania) identified TEM as the most common ESBL type. All the four studies that reported the dominance of CTX-M-15 were published between 2008 and 2012, where as the one study that had reported TEM as the most common ESBL type was published in 2003 (Sonda et al., 2016). In addition to the detection of molecular determinants of ESBL resistance in *E. coli* and *Klebsiella spp.,* a few studies from SSA have also reported detection of molecular determinants of ESBL resistance in *Salmonella enterica* isolates. In Malawi, an incHI2 plasmid which harboured $bla_{\text{CTX-M-15}}$, $bla_{\text{OXA-30}}$ and several other AMR genes was identified in a *S.* Typhimurium strain through whole genome sequencing (WGS) (Feasey et al., 2014a).

1.8.2 Molecular determinants of fluoroquinolone resistance

Characterisation of molecular determinants of fluoroquinolone resistance in SSA indicates that resistance to fluoroquinolone in the region is mediated through multiple mechanisms. These mechanisms include mutations in the *gyrA, parC* and *parE* genes, efflux pumps and plasmid-mediated genes. In Kenya, all 17 *E. coli* isolates with resistance to ciprofloxacin and levofloxacin had amino-acid substitutions on codon positions 83 (Ser83 ->Leu) and 87 (Asp87 -> Asp) in the QRDR of the GyrA amino acid sequence and the plasmid-*borne aac(6')-1b-cr2*. Five

of the isolates also had plasmid-mediated *qnrA* and *qnrB* genes. Two silent mutations were also detected outside the QRDR of the GyrA amino acid sequence on codon positions 156 Proline in all the 17 isolates (Kariuki et al., 2007). The mutations in the QRDR and identified in this study have all been shown to confer fluoroquinolone resistance (Jacoby et al., 2014). It is however not known whether there are synonymous mutations outside QRDR that have any role in fluoroquinolone resistance. High level resistance to fluoroquinolones was also reported in *E. coli* isolates from Nigeria and there, ciprofloxacin resistance was associated with isolates having at least two codon mutations in the GyrA (S83L, D87N) or Par*C* (S80I, Y74C) amino acid sequences and in some isolates the plasmid mediated *qnrS1* gene (Lamikanra et al., 2011).

A number of studies have also investigated molecular determinants of fluoroquinolone resistance in *S. enterica* and results suggest that whilst *S. enterica* isolates from SSA increasingly show reduced susceptibility to fluoroquinolone, full fluoroquinolone resistance remains uncommon (Chattaway et al., 2016). An investigation on a single *S.* Typhi isolate with full resistance to fluoroquinolones in South Africa identified a combination of efflux activity and a single codon mutation (Ser83 to Tyr) on the GyrA amino acid sequence as the mechanism associated with the fluoroquinolone resistance (Keddy et al., 2010). Another study found a combination of mutations in the GyrA and ParC amino acid sequences and efflux pumps to only confer reduced susceptibility rather than high level resistance to ciprofloxacin in *S.* Typhi (Smith et al., 2010). Significant associations of chromosomal mutations and plasmid-mediated genes with reduced susceptibility to

fluoroquinolone in *S*. Typhi were also reported in studies from other countries such as Zambia (Hendriksen et al., 2015), DRC (Lunguya et al., 2012) and Kenya (Al-Emran et al., 2016). *S*. Typhi isolates with DCS in DRC had a single codon mutation at positions 83 (Ser83 to Phe or Tyr [in 22/31 isolates]) or 87 (Asp87-> Gly or Tyr or Asn, [in 9/31 isolates] in the GyrA amino acid sequence (Lunguya et al., 2012). Similar mechanisms were also reported in Zambia where one isolate with codon mutation Aspn87-> Asn and another with Ser83->Tyr in the GyrA amino acid sequence had DCS but full resistance to nalidixic acid (Hendriksen et al., 2015).

1.8.3 ESBL producing and fluoroquinolone resistant clones of *E. coli* and *K. pneumoniae*

Third generation cephalosporin resistance in *E. coli*, which is predominantly through production of the CTX-M-15 ESBL enzymes, is associated with a specific pandemic clone of the sequence type ST131. This ST131 CTX-M-15 clone is also associated with chromosomal mutations that confer fluoroquinolone resistance (Figure 1.6) and consequently, the global dissemination of both the ESBL and fluoroquinolone resistance in *E. coli* has been attributed to its expansion (Petty et al., 2014; Nicolas-Chanoine et al., 2008).

The most well-known and widely disseminated clone of *K. pneumoniae,* ST258, is associated with carbapenem resistance (Bowers et al., 2015). ESBL and fluoroquinolone resistance in *K. pneumoniae* is associated with multiple localised clonal lineages and unlike in *E. coli* the spread of ESBL and other plasmid-mediated

genes is mostly attributed to horizontal gene transfer (Holt et al., 2015; Calbo and Garau, 2015). Some of the most reported *K. pneumoniae* associated with ESBL resistance globally are ST11, ST15, ST14 as well as ST258.

It is still unclear which ESBL producing and fluoroquinolone resistant E. *coli* and *K. pneumoniae* clones are circulating in SSA (Chattaway et al., 2016). CTX-M-15 production has been detected in some *E. coli* ST131 as it has been in other STs but no dominance of a single clone has been reported (Mshana et al., 2016; Ribeiro et al., 2016). ESBL producing *K. pneumoniae* ST14 has also been identified in a number of countries within SSA (Mshana et al., 2013a; Poirel et al., 2011; Jacobson et al., 2015). There are suggestions that ST14 could be the most common ESBL producing *K. pneumoniae* ST in the region (Mshana et al., 2013a) but more detailed studies across different countries are required to ascertain this. As earlier pointed out, CTX-M-15 producing strains especially in *E. coli* also tend to harbour genotypes that confer resistance to fluoroquinolones (Nicolas-Chanoine et al., 2014) and it is possible the emerging CTX-M-15 producing clones in SSA also harbour fluoroquinolone resistance genotypes, though this has not been confirmed.

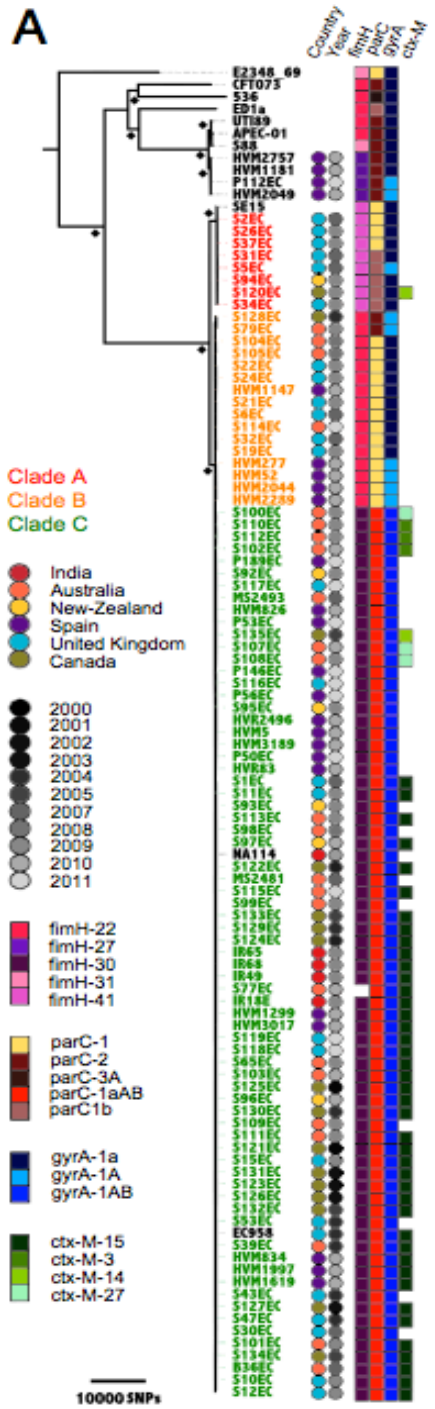

Figure 1.6 Phylogenetic relationships of CTX-M-15 associated *E. coli* ST131 clone in context of other ST131 strains. (A) ML phylogram of ST131 strains coloured by clade (red =clade A, orange=clade B and green=clade C, black=reference strains) (Petty et al., 2014).

1.9 Factors driving the spread of AMR

Multiple factors have been implicated in driving the spread of AMR, and it is clear that use of antimicrobial agents in hospitals, community or agriculture is a major driving factor in the spread of AMR (Holmes et al., 2016). It is very challenging to attribute which aspect of antimicrobial usage is most important, particularly because many of such factors are to do with the general human behaviour, for example, in relation to prescriptions by clinicians or patient adherence to treatment regime. Changing human behaviour towards antimicrobials is challenging as so many factors contribute to their use, including high disease burden, lack of diagnostic facilities to guide prescription, use of antimicrobial agents in agriculture, over prescription by private health care providers and patient demand (Laxminarayan et al., 2013; Holmes et al., 2016). To this end, there is an urgent need for behavioural studies to understand how people understand the nature of their infections and treatment courses.

It is important to note that the evolution of many AMR mechanisms predate the development and usage of antimicrobial agents by humans by millions of years (Holmes et al., 2016). Antimicrobial use does however contribute to the selection pressure that favours the proliferation of resistant over susceptible strains, and also selects for the assembly and evolution of AMR mutations and transfer of MGEs harbouring multiple AMR genes between strains (O'Brien, 2002; Holmes et al., 2016).

Additional factors favouring the emergence of resistance include suboptimal doses

or durations of therapy, and the use of counterfeit or expired antimicrobial agents which might select populations with low level resistance to an antimicrobial and favour the development of high level resistance (Laxminarayan et al., 2013; Hooper and Jacoby, 2015a).

Furthermore, travel and globalisation facilitate spread of AMR within and across regions (Holmes et al., 2016; Harbarth and Samore, 2005). Travellers from places where resistance to an antimicrobial agent has emerged may be colonised with the resistant bacteria and spread it to places where resistance has not emerged while global mixing and high population densities facilitate rapid transmission of DRI (Harbarth and Samore, 2005). Studies on sequential emergence of resistance to some antibiotics in different places have shown a correspondence between emergence of AMR and major travel routes (Figure 1.7).

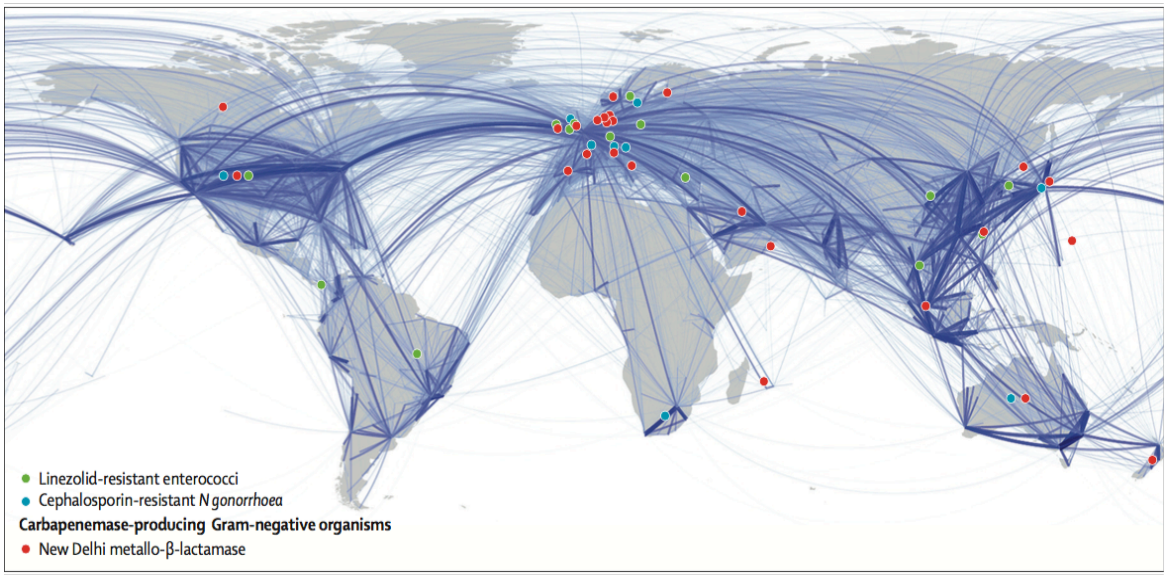

Figure 1.7 World Map showing the relationship between travel routes and emergence of carbapenem, cephalospoirn and Linezolid resistance in Gram-negative bacteria, *N. gonorrhoea* and *Enterococci* respectively (Holmes et al., 2016).

1.10 Concluding remarks and rationale

SSA suffers a high burden of severe bacterial infection, but has a paucity of diagnostic facilities to guide prescription. The considerable need to use antimicrobials in the absence of adequate diagnostics or antimicrobial stewardship is creating a fertile environment for the emergence of AMR. Whilst the trends in the emergence of ESBL and fluoroquinolone resistance appear not be very different from those in other regions on the globe, there has been a limited number of studies describing and characterising AMR and DRI in this region. This has been as a result of lack of resources and limited research capacity in most countries in SSA. As a

result, there are wide gaps in understanding the epidemiology and evolution of AMR in this region.

In SSA, cephalosporins and fluoroquinolones have become the antimicrobial agents of choice, particularly, but not exclusively, for patients admitted to health care facilities for the treatment of severe bacterial infections. Recent data suggest the emergence and increasing spread of ESBL and fluoroquinolone resistance among major pathogens such as *S. enterica*, *E. coli* and *K. pneumoniae* in SSA, but reliable estimates of burden of DRI remain unknown in the majority of countries in the region. Studies describing the global population structure of bacteria responsible for ESBL and fluoroquinolone resistance have often excluded data from Africa. In the absence of such data it is unclear whether the clades associated with ESBL and fluoroquinolone resistance are the same in SSA as globally. However, previous studies on NTS have identified that markedly different clades are circulating in SSA when compared to the rest of the world and these differences could also be true for other key pathogens, such as *E. coli* and *K. pneumoniae* that are mostly associated with ESBL and fluoroquinolone resistance.

Malawi, as is the case with a majority of countries in SSA, is under considerable pressure from poverty, undernutrition, urbanisation, malaria and HIV. These factors have created a conducive environment for the proliferation of SBI, which has been characterised by MDR especially to the first-line antimicrobial agents. It is evident that resistance to antimicrobial agents such as cephalosporins, fluoroquinolone and gentamicin has also emerged in Malawi and other SSA countries, posing a threat of

untreatable infections in settings where these are the last resort antimicrobial agents.

The CDC identified tracking and characterisation of resistant organisms and identifying drivers of the emergence and spread of AMR to be key to limiting the emergence and further spread of resistant bacteria (Interagency Task Force on Antimicrobial Resistance, 2012).Elsewhere, studies that utilise surveillance, WGS and mathematical modelling approaches have been useful in detecting and characterising resistant bacteria and modelling the dynamics of DRI (Spicknall et al., 2013; Petty et al., 2014; Holt et al., 2015; WHO, 2014). However, with such studies being rare in Malawi, especially with regard to bacteria associated with ESBL and fluoroquinolone resistance, the full burden of ESBL and fluoroquinolone resistance, the molecular mechanisms and clones associated with the ESBL and fluoroquinolone resistance and the key drivers of DRI remain unknown.

1.11 Hypothesis and Aims

1.11.1 Hypothesis

E. coli and *K. pneumoniae* are key Gram-negative pathogens associated with ESBL and fluoroquinolone resistance globally, and the dissemination of ESBL and fluoroquinolone resistance around the world has been associated with a single dominant global lineage of *E. coli* and localised clonal lineages of *K. pneumoniae*. In Malawi *E. coli* and *Klebsiella spp*. expressing ESBL and fluoroquinolone resistance have been detected for over a decade. It is here hypothesised that:

- There has been high and increasing burden of DRI in Blantyre in all pathogens and to all drug classes in Malawi;
- The dissemination of ESBL and fluoroquinolone resistance has been as a result of expansion of specific distinct clones *E. coli* and *K. pneumoniae* lineages and not HGT.

1.11.2 Aims

- Describe the trends in incidence of BSI and prevalence of AMR in bacterial pathogens in Blantyre
- Identify the molecular determinants of ESBL and fluoroquinolone resistance in *E. coli* and *Klebsiella spp*.
- Describe the population structure of *E. coli* and *Klebsiella spp.* and identify the *E. coli* and *K. pneumoniae* lineages associated with ESBL and fluoroquinolone resistance in Blantyre.
- Develop mathematical models useful for understanding the dynamics of DRI.

Chapter Two: Methods and Materials

2.1 Overview

This PhD project was undertaken across three sites: MLW, Liverpool-School of Tropical Medicine (LSTM), Liverpool, United Kingdom (UK) and WTSI. Methodologically, the project had three components:

1. Retrospective analysis of laboratory surveillance data
2. Phylogenetic analysis of whole genome sequence data to define population structure and identify antimicrobial resistance genes
3. Mathematical modelling of incidence of BSI in a population with heterogeneous susceptibility

Blood culture surveillance was undertaken by the MLW diagnostic laboratory in Malawi and whole genome sequencing at WTSI, UK. All analyses were carried out at MLW, LSTM and WTSI.

2.2 Retrospective study of community acquired bloodstream infection and antimicrobial resistance in Blantyre

2.2.1 Study setting

Malawi is landlocked country located in southeast Africa, sharing her borders with

Zambia to the northwest, Tanzania northeast, and Mozambique to the south, southeast and southwest. Malawi's total land area is 118,484 km^2 of which 24,404 km^2 (20.6%) is water (Figure 2.1). As of 2015, Malawi's population was estimated at 17.2 million (World Bank, 2017c) with an estimated annual growth rate of 3.32% (Central Intelligence Agancy, 2017).

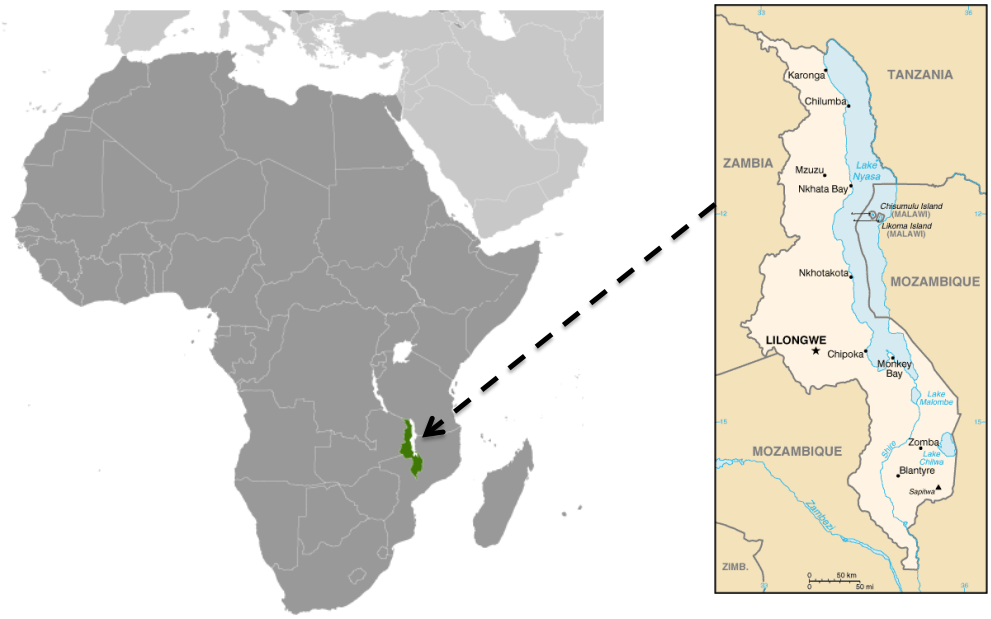

Figure 2.1 Map of Africa and Malawi (Central Intelligence Agancy, 2017)

Malawi is ranked amongst the top ten poorest countries in the world. The economy is heavily dependent on agriculture and shows very inconsistent patterns. In 2011 70.9% of the population lived below the international standard poverty line (< $1.90 per day) (World Bank, 2017b). The current proportion of the population living below the poverty line is likely to be higher as growth domestic product

(GDP) per capita has declined from $524.3 in 2011 to $372.0 in 2015 (World Bank, 2017a)

HIV prevalence is high in Malawi but substantial progress has been made in reducing HIV transmission and provision of ART to infected individuals. Overall, HIV prevalence rates have declined from 16.4% in 1999 to 8.8% in 2016 but urban areas have higher rates, twice as high as rural areas (National AIDS Commission, 2015; National Statistical Office (NSO) and ICF, 2017). Enrolment on the ART programme in Malawi has increased from 4,000 (2.3% of those in need of ART) in 2004 to over 530,000 (67.0% those in need of ART) in 2014 (National AIDS Commission, 2015). Malawi also suffers a high burden of other infectious diseases, particularly bacterial infections and malaria.

Blantyre is one of Malawi's two principal cities and has a rapidly expanding population. Together with the surrounding peri-urban rural areas of Blantyre district, the population of Blantyre was estimated at 1.3 million people in 2016. The population for Blantyre city alone was estimated at 920,000 in 2016 (National statistical Office (NSO)). HIV prevalence was estimated at 17.6% in 2016 from 22.3% in 2004 (National Statistical Office (NSO) and ORC Macro, 2005; National Statistical Office (NSO) and ICF, 2017).

2.2.1.1 Queen Elizabeth Central Hospital

QECH is one of the largest government hospitals in Malawi and is the only public hospital providing free medical care to Blantyre's urban population. There are

approximately 1,000 beds, though occupancy frequently exceeds capacity. The hospital admits approximately 10,000 adult (aged ≥ 16 years) and 50,000 paediatric (aged < 16 years) medical patients per year (Feasey et al., 2015c).

2.2.1.2 Malawi-Liverpool- Wellcome Trust Clinical Research Programme

Located within the QECH campus, MLW is one of the five major overseas programmes for the Wellcome Trust and was established in 1995 as an affiliate of the College of Medicine (University of Malawi) in partnership with LSTM and the University of Liverpool. It is a biomedical and clinical research centre of excellence, which conducts research relevant to health in SSA (www.mlw.medcol.mw).

2.2.2 Microbiology procedures

Since 1996 MLW has been conducting bacteraemia surveillance by blood culture at QECH, and this surveillance became a routine clinical service in 1998. All Malawian patient blood culture isolates described in this thesis are from blood samples taken as part of the routine surveillance by the clinical staff at QECH. Laboratory processing of blood samples was performed at the MLW Diagnostic Laboratories.

Seven to ten mls of blood were taken for culture under aseptic conditions from all adult patients admitted to the QECH with fever (axillary temperature > 37.5°c) or clinical suspicion of sepsis and inoculated into a single aerobic bottle (BacT/Alert, bioMérieux Marcy-L'Etoile, France). Two to five mls of blood were also taken from children with febrile illness, but with no malaria and no obvious focus of infection,

or severely ill with sepsis and patients who failed initial malaria treatment but remained febrile (Everett et al., 2011). Afebrile patients were unlikely to have blood samples for culture unless critically ill and sepsis was suspected.

Blood culture had been undertaken using the automated BacT/ALERT system (BioMérieux, France) since 2000. Prior to 2000 manual culture was performed as previously described (Gordon et al., 2001). Enterobacteriaceae and oxidase positive Gram-negative bacilli were identified by analytical profile index ([API], BioMérieux, France). *Salmonellae* were further serotyped by Lancefield antigen testing according to the White-Kauffmann-Le Minor scheme by the following antisera; polyvalent O & H, O4, O9, Hd, Hg, Hi, Hm, and Vi antisera (Pro-Lab Diagnostics, UK). *Staphylococci* were identified by tube coagulase and ß-haemolytic *Streptococci* identified by Lancefield antigen testing. Bacteria that form part of the normal skin or oral flora including Diphtheroids, *Bacilli*, *Micrococci*, coagulase negative *Staphylococci*, and alpha-haemolytic *Streptococci* (other than *S. pneumoniae*) were considered to be contaminants, even though they may act as pathogens under specific situations such as in HIV infected individuals (Barrow, 1993).

2.2.3 Antimicrobial susceptibility testing

Antimicrobial susceptibility tests were performed by the disc diffusion method following the British Society of Antimicrobial Chemotherapy (BSAC) methods and breakpoints guidelines (www.bsac.org.uk). Testing was routinely limited to one plate containing six discs of ampicillin (or penicillin for Gram-positive pathogens),

chloramphenicol, cotrimoxazole, ceftriaxone, ciprofloxacin and gentamicin. However, some isolates especially Gram-positive isolates were also tested for susceptibility to erythromycin and tetracycline. Methicillin resistance in *S. aureus* was inferred by cefoxitin resistance, which replaced methicillin resistance testing in 2010.

2.2.3.1Screening for production of ESBLs

Isolates were screened for ESBL-producing status using a cefpodoxime (10μg) disc since 2007, prior to which a ceftriaxone disc was used. Isolates that exhibited sensitivity (BSAC zone size ≥ 26) to cefpodoxime were recorded as ceftriaxone sensitive. Isolates that expressed resistance (BSAC zone size ≤ 25) to cefpodoxime by the disc diffusion method had a confirmatory E-test done to determine presence or absence of ESBLs. The E-test had ceftazidime on one end of the strip and ceftazidime plus clavulanic acid on the other end. The MIC of each end of the strip were determined and the ratio of ceftazidime MIC to ceftazidime plus clavulanic acid MIC was calculated as follows:

$$\frac{Ceftazidime\ (TZ)MIC}{Ceftazidime + Clavulanic\ acid\ (TZL)MIC}$$

An isolate was considered to be ESBL producing if the above ratio was ≥ 8; otherwise it was considered as non-ESBL producing.

2.2.4 Data management and statistical analyses of surveillance

Blood culture data for analysis were obtained from MLW's Laboratory Information Management System (LIMS) database. Since 2010, data were entered directly into the LIMS database, however prior to that, blood culture data were transcribed from specimen request forms, which were also used as laboratory work sheets onto log-books and these were eventually transferred via double entry into an electronic database. The surveillance data transferred from the books to the electronic LIMS were previously validated in Dr. N.A. Feasey's PhD project (Feasey, 2014). Further validation and cleaning has also been undertaken in this thesis. Only data for blood culture isolates sampled between 1998-2016 were analysed. Isolates were identified by unique identification (ID) numbers of the patients from whom the blood culture sample was taken. All records with duplicate ID numbers were excluded.

Descriptive statistics were used to summarise distribution of blood cultures, prevalence of pathogens (in absolute numbers) over time and age of patients.

Ampicillin, chloramphenicol and cotrimoxazole were categorised as first line antimicrobial agents for Gram-negative pathogens whereas penicillin, cotrimoxazole and chloramphenicol were categorised as first-line antimicrobial agents for Gram-positive pathogens. Isolates expressing phenotypic co-resistance to ampicillin (or penicillin for Gram-positive pathogens), cotrimoxazole and chloramphenicol were referred to as resistant to first-line (RFL). Ceftriaxone, ciprofloxacin and gentamicin were categorised as second-line antimicrobial agents. Isolates expressing

phenotypic resistance to at least three antimicrobial agents (when both first-line and second line agents were considered) were categorised as MDR. Proportions of isolates resistant to a particular antimicrobial agent were obtained by dividing the number of isolates identified to be resistant to the agent by the total number of isolates tested for susceptibility to that agent.

Trends in proportions of isolates resistant to an antimicrobial agent over time were shown in tables and graphs and tested for significance using the Cochran-Armitage trend test. Independent pairwise comparisons of AMR proportions where done using Chi-square test for proportions; where a chi-square test was not appropriate (i.e. where cell counts of ≤ 5 were observed), Fisher's exact test was used. All analyses were done in R (Version 3.1.3) Statistical Package (R Core Team 2014).

2.2.4.1 Calculation of incidence rates

Population denominators

Population denominators for 1998-2007 were obtained from the population estimates for Blantyre city from the 1998-2023 Malawi population projections by the National Statistical Office (NSO). 2008-2014 population denominators were obtained from NSO's 2008-2030 national population projections updated from the 2008 population census (National statistical Office [NSO]). Age stratified estimates were only available for the 2008-2014. In this thesis, mean ratios of age group population sizes for 2008-2014 (Table 2.1B) were calculated and used to split the

overall population size estimates of 1998-2007 into five-year age-stratified population sizes.

Table 2.1A Age stratified population denominators for Blantyre City (1998-2007)

Age	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007
0-4	87787	88287	87620	95116	95116	96616	99281	102113	105278	107443
5-9	68633	69024	68503	74364	74364	75536	77620	79834	82308	84001
10-14	56429	56750	56322	61140	61140	62104	63817	65638	67672	69064
15-19	53715	54020	53613	58199	58199	59117	60747	62480	64417	65742
20-24	59431	59769	59318	64393	64393	65408	67212	69130	71272	72738
25-29	58405	58737	58294	63281	63281	64279	66052	67936	70041	71482
30-34	46301	46565	46214	50167	50167	50958	52364	53857	55527	56669
35-39	31829	32010	31768	34486	34486	35030	35996	37023	38171	38956
40-44	20906	21025	20867	22652	22652	23009	23644	24318	25072	25588
45-49	14116	14197	14089	15295	15295	15536	15965	16420	16929	17277
50-54	10234	10292	10214	11088	11088	11263	11573	11903	12272	12525
55-59	7103	7144	7090	7696	7696	7818	8033	8262	8518	8694
60-64	4767	4794	4758	5165	5165	5247	5391	5545	5717	5835
65-69	3153	3171	3147	3416	3416	3470	3566	3667	3781	3859
70-74	1940	1951	1936	2101	2101	2135	2193	2256	2326	2374
75-79	1158	1165	1156	1255	1255	1275	1310	1347	1389	1418
≥ 80	1092	1098	1090	1183	1183	1202	1235	1270	1310	1337
Total ('000)	**527**	**530**	**526**	**571**	**571**	**580**	**596**	**613**	**632**	**645**

Table 2.1B Age stratified population denominators for Blantyre (2008-2014)

Age	2008	2009	2010	2011	2012	2013	2014	2015	2016
0-4	103967	111942	119619	126962	133922	140379	144073	147747	151328
5-9	85013	88740	92722	97087	101913	107292	115343	123145	130648
10-14	72466	74082	76639	79876	83469	87191	91037	95135	99614
15-19	77248	76449	75697	75228	75327	76200	78004	80744	84152
20-24	83094	84410	85015	85084	84857	84514	84133	83794	83736
25-29	74028	77646	81038	84144	86893	89267	91071	92174	92751
30-34	53625	57709	62101	66622	71046	75193	79082	82771	86187
35-39	36183	38989	41994	45232	48749	52563	56721	61198	65807
40-44	23870	25510	27425	29601	32015	34634	37449	40465	43714
45-49	17435	18159	18942	19851	20955	22311	23939	25835	27986
50-54	12656	13314	13970	14625	15289	15980	16700	17483	18377
55-59	8332	8878	9502	10174	10853	11502	12134	12762	13397
60-64	5925	6198	6453	6722	7049	7464	7974	8553	9179
65-69	3676	3923	4217	4533	4840	5123	5375	5617	5868
70-74	2482	2560	2630	2713	2832	2995	3207	3455	3726
75-79	1352	1458	1574	1688	1779	1858	1923	1982	2054
≥ 80	1637	1568	1522	1501	1509	1535	1577	1636	1702
	662989	**691535**	**721060**	**751643**	**783297**	**816001**	**849742**	**884497**	**920226**

Annual incidence rates (per 100,000) of BSI by a pathogen were estimated by dividing the number of BSI cases observed in that year by the estimated mid-year population size of Blantyre City and multiplying this ratio by 100,000

i.e.:

$$I = \frac{n_{ij}}{P_j} \times 100000$$

Where:

$I =$ incidence; $n_i =$ number of BSI cases by pathogen i in year j;

$P_j =$ Estimated mid year population size for Blantyre city in year j.

Incidence was also calculated for five-year age stratified groups by dividing the total number of BSI cases by person years of observation for the population denominators stratified by age and multiplied by100,000.

i.e.:

$$I_g = \frac{n_{ig}}{P_g} \times 100000$$

Where $I_g = is\ the\ incidence\ of\ by\ a\ pathogen\ in\ age\ group\ g$;

$n_{ig} = total\ number\ of\ BSI\ by\ pathogen\ i\ in\ age\ group\ g$.

$P_{g=}$age stratified person years of observation for the population denominators

Trends in estimated minimum incidence of BSI by high frequency and medium frequency pathogens were modelled over time and tested for significance using the Poisson regression model as follows:

$$\ln(\mu/P) = \beta_0 + \beta_1 * \ln(t)$$

which becomes

$$\ln(\mu) = \beta_0 + \beta_1 * \ln(t) + \ln(P)$$

where μ is mean of number of BSI cases;

$t\ is\ time$ ($in\ years$);

P is the mid-year population size;

$\beta_0$and β_1 are constant coefficients.

Whenever there was over dispersion with the Poisson regression, the Negative Binomial regression model, which is a generalisation of the Poisson regression model, was instead used. Unlike the Poisson regression model, the negative binomial model does not assume equality of the mean and variance.

2.3 Whole genome sequencing of *E. coli* and *K. pneumoniae* isolates

2.3.1 Study isolates

Isolates selected for sequencing were collected as part of the routine bacteraemia and meningitis surveillance undertaken at QECH by MLW. Further isolates from rectal swabs were collected as part of a study aimed at identifying ESBL producing isolates in hospital carriage isolates. The bacteraemia surveillance, from which the blood culture isolates were obtained, has been described in section 2.2 of this chapter. The meningitis surveillance and the collection of rectal swab isolates are described below.

2.3.1.1Meningitis surveillance

In addition to the blood culture surveillance (section 2.2), MLW also instituted meningitis surveillance at QECH in 2000. Briefly, under the meningitis surveillance, adult and paediatric patients who presented to the hospital with clinical suspicion of meningitis underwent lumbar puncture where 5-10mls of cerebrospinal fluid (CSF)

was collected from adults and 1-2mls of CSF was collected from children. Gram-stain was performed for culture with CSF white cell count > 10 cells/μL. Adult CSF was stained with India ink to detect and separate fungal meningitis samples from bacterial meningitis samples. All samples were cultured on sheep blood agar (SBA) and chocolate agar for 48 hours under aerobic and microaerophilic conditions (Wall et al., 2014).

2.3.1.2 Rectal swab isolates

Rectal swab isolates were collected over a period of two weeks at QECH. Samples were obtained from patients using a sterile swab. Swabs were spread on Maconkey agar with ceftriaxone and incubated at 37˚C overnight. Resulting colonies were identified biochemically using API 20E test (Biomerieux).

2.3.1.3 Isolate selection

Isolates were selected with the aim of understanding the diversity of the *E. coli* and *K. pneumoniae* populations in Blantyre in relation to AMR, hence to maximise this diversity, selection was based on isolates' phenotypic resistance profile, year of isolation and clinical source of isolation. Antimicrobial susceptibility tests were done by the disc diffusion test (Section 2.2.3) to determine the phenotypic resistance profiles to six commonly used antimicrobial agents at QECH including ampicillin, chloramphenicol, cotrimoxazole, ceftriaxone, ciprofloxacin and gentamicin.

2.3.1.4 Other global isolates included for analysis in this thesis

In order to place the Malawian *E. coli* and *K. pneumoniae* isolates into a global

context and determine if the clones associated with ESBL and fluoroquinolone resistance globally are the same as those in Malawi, isolates from previously published global population structure studies of *E. coli* (Petty et al., 2014) and *K. pneumoniae* (Holt et al., 2015) were selected (stratified by phylogenetic clustering). Raw read sequence data for the selected isolates were retrieved from the European Nucleotide Archive (ENA) under the study numbers ERP001354, ERP004358 (Petty et al., 2014) and RP000165 (Holt et al., 2015) and analysed together with the sequenced data of the Malawian isolates (see the following sections on how Malawian isolates were sequenced).

2.3.2 Processing of genomic DNA

2.3.2.1 Culturing of selected E. coli and K. pneumoniae isolates

Samples were stored in Microbank bacterial pre-servers (Prolab Diagnostics, Ontario, Canada) at -80°C. *E. coli* and *K. pneumoniae* isolates selected for culture were retrieved from the freezers, streaked onto SBA plates, and incubated at 37ºC for 18-24 hours in 5% carbon dioxide (CO_2) incubator. Whenever there was bacteria growth on the primary plate, a purity plate was made by sub-culturing a single colony of *E. coli* or *K. pneumoniae* to a fresh plate of SBA incubated at 37ºC for 18 -24 hours in 5% CO_2 incubator. A single colony from the purity plate was inoculated in 10ml of sterile Luria Broth (LB) and incubated overnight.

2.3.2.2 Genomic DNA extraction

100µl of DNA was extracted from 300µl of LB broth using the QIAamp One for All

Nucleic Acid Kit (Qiagen, Hilden, Germany) on the Qiagen Biorobot Univeral System (Qiagen, Hilden, Germany) according to manufacturers instructions. One µL of extracted DNA was run on gel electrophoresis with a 100 bp DNA ladder to approximate levels of DNA concentration.

The first 50µl of DNA was shipped by courier to WTSI in capped 1.5 ml tubes on dry ice, in tightly sealed containers. The remaining 50 µl was stored at -20C, to be used in case of loss during shipping.

2.3.3 Genomic DNA sequencing

All the Malawian *E. coli* and *K. pneumoniae* isolates were sequenced by the Core Sequencing Teams at WTSI. Multiplexed genomic DNA libraries were prepared and sequenced using the Illumina Hiseq 2000 platform (Illumina, Inc., San Diego, California). This generated paired-end sequence reads of 100bp in length. The sequence reads were trimmed and checked for quality using an average base pair quality phred score of at least 20 per read as the minimum quality score threshold. Short read sequences were deposited in ENA study number PRJEB8265.

2.4 Bioinformatics analyses

2.4.1 Quality checking and control for sequence data

Sequenced read data went through an automated quality checking and control that produces base quality statistics and does adapter trimming. Both *E. coli* and *K.*

pneumoniae have average genome size of five mega base pair (MB). In addition to the automated quality checking and control, isolates with genome sizes substantially smaller or larger than 5MB were also removed from further analysis.

2.4.1.1In silico classification of sequenced strains

Kraken, a program which uses k-mer and Least Common Ancestor database to assign DNA sequences to a matching organism (Wood and Salzberg, 2014), was used to confirm the classification of the *E. coli* and *K. pneumoniae* isolates sequences to species level. While its accuracy is comparable to that of BLAST (Altschul et al., 1990), one of the best known methods for taxonomic classification of unknown sequences, Kraken is faster than BLAST and other comparable programs such as MEGAN or Naïve Bayes Classifier (Wood and Salzberg, 2014).

Kraken was run on default settings (k-mer size, k=31) for both the *E. coli* and *K. pneumoniae* whole genome sequence datasets. Genome sequences that did not match either *E. coli* or *K. pneumoniae* genomes sequences were excluded from further analysis. However, when genome sequences from a DNA sample initially classified as *E. coli* matched a *K. pneumoniae* genome sequence then the sample was reassigned to *K. pneumoniae* genome dataset for further analysis. Similarly, genome sequences of samples initially identified as *K. pneumoniae* were reassigned and analysed as *E. coli* genome sequences if Kraken classified them as *E. coli*. These results were also corroborated with results of mapping the genomes sequences to their respective reference genomes as described in the following section.

2.4.2 Mapping of genome read sequences to reference genomes sequences

Sequence reads of confirmed *E. coli* genomes were initially mapped to the *E. coli* strain K-12 sub-strain MG1655 reference whole genome sequence (accession number U00096) using SMALT. SMALT is a tool that aligns DNA read sequences with genomic sequence references using a combination of hashing and dynamic programming. The choice of SMALT for mapping sequence reads to a reference genome over other popular Burrows-Wheeler Transformation (BWT) algorithm based aligners such as the Burrows-Wheeler Alignment tool (BWA) (Li and Durbin, 2009) was based on the fact that SMALT is more tolerant to high genomic variations, which are common in bacteria genomes, particularly *E. coli* and *Klebsiella spp.*, than the BWT based tools (Caboche et al., 2014).

SMALT was run with the option of randomly mapping repeats. The mapping of the *E. coli* sequence reads to the reference genome generated binary alignment map (BAM) files. Realignment of the insertion and deletion (indel) sites in the generated BAM files was performed using GATK v3.3.0 (McKenna et al., 2010). Consensus pseudo-sequences were generated from each BAM file and aligned with other sequences to generate whole genome alignment of the study isolates.

Following the same procedure, sequence reads of *K. pneumoniae* genomes were mapped to the *K. pneumoniae* MG78578 (accession number PRJNA31) reference genome.

2.4.3 Genome sequence assembly and annotation

The *E. coli* and *K. pneumoniae* genome read sequences were also put on an automated assembly and annotation pipeline to generate contiguous sequences (contigs) and genome annotation files. Assembly of sequence was done using Velvet v1.2.09, (Zerbino and Birney, 2008) by selecting the most optimal k-mers using Velvetoptimiser (Zerbino, 2010). Velvet is a set of algorithms that uses the De Bruijn graph to assemble short read sequences *de novo*. Assembled contigs with less than 300bp were excluded from the dataset and the remaining contigs were scaffolded using SSPACE Basic v2.0 (Boetzer et al., 2011). The gaps between contigs were filled by GapFiller v1.10 (Nadalin et al., 2012). Sequence assemblies were annotated using Prokka v1.11 bacterial annotation tool (Seemann, 2014).

2.4.4 *In silico* multi-locus typing

Multi-locus sequence typing (MLST) was first proposed in 1998 as an unambiguous and portable method for characterising bacteria using nucleotide sequence data from house-keeping genes (Maiden et al., 1998). MLST provides a unified approach of identifying isolates based on the sequence of each gene fragment, which can be compared in the form of an allelic profile (Maiden et al., 1998). Additionally, given a set of isolates from a bacterial species with an MLST scheme, MLST generates data that can also be used to for evolutionary and population studies independent of the isolates' genetic diversity, population structure or evolution (Maiden, 2006). The reduced speed and cost of nucleotide determination following the advent of next-generation sequencing and the availability of improved web-based databases and

analysis tools have seen MLST become one of the most important and commonly used methods for characterising bacterial strains.

Here, *in silico* MLST was performed by a BLAST search of sequence assemblies against the PubMLST database using an in-house WTSI script get_sequence_type.py to determine the sequence types (STs) of the Malawian *E. coli* and *K. pneumoniae* isolates.

For *E. coli*, different allelic profiles of each isolate were determined based on the seven house keeping genes described in Table 2.2. For each isolate, the allelic profile was compared against the *E. coli* MLST database containing all known *E. coli* MLST allelic profiles and their corresponding STs. Once a match was identified, the corresponding ST was assigned to the isolate.

Table 2.2 Seven house-keeping genes for *E. coli* MLST

Gene notation	**Encoded protein**
adk	Adenylate kinase
fumC	Fumurate hydratase
gyrB	DNA gyrase
icd	Isocitrate/isopropylmalatedehydrogenase
mdh	Malate dehydrogenase
purA	adenylosuccinate dehydrogenase
recA	ATP/GTP binding motif

Similarly, MLST allelic profiles were determined for each of the *K. pneumoniae* strains based on the seven *K. pneumoniae* house-keeping genes described in Table 2.3 and the corresponding STs assigned to the isolates.

Table 2.3 Seven housekeeping genes for Klebsiella MLST scheme

Gene notation	Encoded protein
rpoB	Beta-subunit of RNA polymerase
gapA	glyceraldehyde 3-phosphate dehydrogenase
mdh	Malate dehydrogenase
pgi	Phosphoglucose isomerase
phoE	phosphorine E
infB	translation initiation factor 2
tonB	periplasmic energy transducer

For either the *E. coli* or *Klebsiella* collections, whenever a new ST was found (not already in the database), the get_sequence_type script assigned the isolate to nearest ST in the database and flagged it as an "Unknown". If however, the allelic profile was a novel combination (and not close to any of the available profiles) the new ST was assigned to the isolate and flagged as "novel".

2.4.5 *In silico* determination of *E. coli* phylogroups

To type the *E. coli* isolates by phylogroup, a two stage in *silico* quadruplex PCR was performed based on the method proposed by Clemont, *et al.* (Clermont et al., 2013). WTSI in-house python script *primers_to_tab.py* was used to extract DNA fragments (primer sequences) of six genes including *chuA, yjaA, arpA, TsPE4.C2, arpAgpE and trpAgpC*, Table (2.1), whenever they were present in a genome sequence.

Table 2.4 Primer sequences for *E. coli* phylogroup typing

Primer ID	Forward sequence	Reverse sequence		
chuA	ATGGTACCGGACGAACCAAC	TGCCGCCAGTACCAAAGACA	**290**	**286**
yjaA	CAAACGTGAAGTGTCAGGAG	AATGCGTTCCTCAACCTGTG	**213**	**209**
TspE4.C2	CACTATTCGTAAGGTCATCC	AGTTTATCGCTGCGGGTCGC	**154**	**150**
arpA	AACGCTATTCGCCAGCTTGC	TCTCCCCATACCGTACGCTA	**402**	**398**
arpAgpE	GATTCCATCTTGTCAAAATATGCC	GAAAAGAAAAAGAATTCCCAAGAG	**303**	**299**
trpAgpC	AGTTTTATGCCCAGTGCGAG	TCTGCGCCGGTCACGCCC	**221**	**217**

The script primers_to_tab.py was run with the option of zero allowed mismatches.

E. coli isolates were firstly categorised into phylogroups based on their allelic profile of the first four gene primers including *arpA, chuA, yjaA*, and *TspE4.C2* as follows: phylogroup A if the isolate was positive for arpA only, phylogroup B1 if positive for *arpA* and *TspE.C2* only, phylogroup F if positive for *chuA* only, B2 if positive for *chuA* and *yjaA* only or *chuA* and *TspE4.C2* only, or *chuA* and *TspE4.C2* only (Table 2.4).

If the isolate's allelic profile for the four primers was not enough to conclusively determine its phylogroup a second stage *in silico* PCR, which in addition to the four previous gene primer sequences (of *arpA*, *chuA*, *yjaA* and *TspE4.C2*), also included primer sequences for *arpAgpC* and *trpAgpC*.

In this second stage, an isolate was assigned to phylogroup A if it were positive for arpA and yjaA only, phylogroup C if positive for *arpA, yjaA* and *trpAgpC*, phylogroup D if positive for *arp* and *chuaA* or *arpA, chuA* and *TspE4.C2* only, and phylogroup E if positive for *arpA, chuA, TspE4.C2* and *arpAgpC* only and phylogroup. Regardless of

an isolate's profile for *trpAgpC* or *arpAgpC,* its phylogroup could not be resolved if it were negative for *arpA*, *chuA, yjaA* and *TspE4.C2*. Table 2.4 summarises how the phylogroups were assigned based on the presence or absence of the gene primer sequences.

Table 2.5 Classification of *E. coli* isolates into phylogroups based on the presence or absence of gene prime sequences

arpA	*chuA*	*yjaA*	*TspE4.C2*	*trpAgpC*	*arpAgpC*	Phylogroup
1	0	0	0	N/A	N/A	A
1	0	0	1	N/A	N/A	B1
0	1	0	0	N/A	N/A	F
0	1	1	0	N/A	N/A	B2
0	1	1	1	N/A	N/A	B2
0	1	0	1	N/A	N/A	B2
1	0	1	0	0	N/A	A
1	0	1	0	1	N/A	C
1	1	0	0	N/A	0	D
1	1	0	0	N/A	1	E
1	1	0	1	N/A	0	D
0	0	0	0	N/A	N/A	Unknown

present =1, absent=0, not applicable=NA

2.4.6 Pan genome and core genome analyses

Roary, a tool that rapidly builds large-scale pan genomes (union of all non-orthologous genes), identifying the core and accessory genes (Page et al., 2015) was used to determine the pan genome and core genome from the annotated genome assembly sequences of the *E. coli* isolates. Ideally, core genome consists of genes conserved in 100% of isolates. However, to allow for random sequencing or

assembly errors, non-orthologous genes were defined as “core” if they were conserved in at least 99% of the isolates. All genes present in less than 99% of the isolates were classified as “accessory”.

Using the same parameters as those used for determining the pan genome and core genome of the *E. coli* isolates, Roary was also used to construct the pan genome and core genome of the *K. pneumoniae* isolates.

2.4.7 Inference of population structure

Recombination events contribute substantially to evolution of many bacteria and may confound phylogenetic reconstruction (Marttinen et al., 2012). In order to remove effects of recombination in inferring population structure of the *E. coli* and *K. pneumoniae* strains and ensure that the topology of their phylogeny is only defined by variations resulting from random mutations, Gubbins was used to remove recombination sites from the multiple sequence alignments obtained from the mapping of short sequence reads to reference sequences. Gubbins is an algorithm that detects and removes regions of elevated sequence diversity in large bacterial genomic datasets to increase the accuracy of phylogenetic reconstructions (Croucher et al., 2015).

An alignment of polymorphic (variable) sites only was generated from the resulting recombination-free whole genome multiple alignments using Snp-Sites (Page et al., 2016). The isolates from either set of *E. coli* and *K. pneumoniae* were then separately clustered into unique subpopulations or sequence clusters (SCs) using hierBAPS

module in the Bayesian analysis of population structure (BAPS) v6.0 software (Corander and Tang, 2007).

2.4.8 Phylogeny reconstruction

Both *E. coli* and *K. pneumoniae* are very diverse species such that it is difficult to find a single reference genome that well represents each of these species. Considering that both the *E. coli* and *K. pneumoniae* isolates in this thesis were selected with the aim of maximising diversity and that as a result high sequence diversity was expected, *E. coli* and *K. pneumoniae* phylogenies were first constructed from the core genome alignment as follows: Orthologous core genes as identified in section 2.4.6 of this chapter were aligned. The individual core gene alignments were concatenated to generate single core genome sequence alignments each for *E. coli* and *K. pneumoniae* that were used to construct a maximum likelihood (ML) phylogenetic tree. The ML phylogenetic tree was constructed using RAxML (Randomised Axelerated Maximum Likelihood) v.7.8.6 (Stamatakis, 2014), which was run under the General Time Reversible (GTR) substitution model with a GAMMA rate of correction heterogeneity.

Bootstrapping was used to assess the reliability of the inferred branches and branch partitions in the phylogenetic trees. Bootstrapping is a method for testing reliability of phylogenetic trees, where in the case of DNA sequences, vertical columns (a set of nucleotides on the same position in different sequences) are randomly selected to produce a sequence alignment of the same length as the original sequences. A

phylogenetic tree is then constructed using the same approach as the one used to construct the phylogenetic tree for the original alignment. This process is repeated for a given number of times and the resulting tree topologies compared. The number of times that the same branch is predicted in a given number of trees constructed is called bootstrap support value. The higher the bootstrap support value the more reliable the branch prediction. In this thesis 100 bootstrap replicates were allowed when reconstructing phylogenies of both the *E. coli* and *K. pneumoniae* genomes.

Whilst core genome alignments were used to reconstruct phylogenies of all isolates in each of the *E. coli* and *K. pneumoniae* collections, polymorphic sites alignments generated from recombination free mapping to reference alignments were used to reconstruct lineage specific phylogenies. In particular, Lineage specific analyses were done with the Bayesian Evolutionary Analysis by Sampling (BEAST) software (Drummond and Rambaut, 2007) which was used to estimate the emergence and evolution of dominant lineage(s).

Tree topologies of core genome phylogenetic trees and phylogenetic trees from polymorphic sites alignments were compared using Dendroscope (Huson and Scornavacca, 2012).

2.4.2.1 Annotation and visualisation of phylogenetic trees

Phylogenetic trees were annotated using the Interactive Tree of Life (iTOL) (Letunic and Bork, 2016) and iCANDY. iTOL was also used for phylogenetic tree visualisation

in addition to another phylogenetic tree visualisation Programme FigTree v1.4 (Rambaut, 2014)

2.4.9 Identification of AMR molecular determinants

Acquired AMR genes were searched for by BLAST v.2.2.30 (Altschul et al., 1990), in a database of acquired AMR genes curated at WTSI based on the Resfinder database (Zankari et al., 2012). Presence of a gene in an isolate was confirmed if its assembled genome sequence had at least 95% nucleotide identity match with a gene in the database and a coverage of at least 90% of the length of the database match.

This method however, was only able to identify acquired resistance genes but not AMR conferring mutations. In order to identify specific amino acid mutations associated with fluoroquinolone resistance, translated nucleotide sequence alignments of the *gyrA, gyrB, parC* and *parE* genes were analysed.

2.4.10 Plasmid replicon typing

It was not possible to reconstruct whole plasmid sequences from the whole genome short read sequence data. However to get an understanding of the nature of plasmids harboured by the *E. coli* and *K. pneumoniae* strains under investigation, some of which carried AMR genes, a plasmid replicon type search against the PlasmidFinder database (Carattoli et al., 2014) was done. A plasmid replicon gene was considered present in a genome if there was a 95% nucleotide identity match in at least 90% gene sequence length.

2.4.11 Statistical analyses

Proportions were compared using Fisher's exact or chi-square tests where appropriate and means of pairwise SNP differences for identified clades were compared using one-way analysis of variance (ANOVA).

2.4.11.1 Modelling the pan genome expansion

In addition to identifying the pan genomes and core genomes, genome samples with sizes of g=1,2,…,N (N=total number of genomes included in the analysis for each collection of either *E. coli* or *K. pneumoniae*) were randomly and iteratively sampled ten times and the pan and core genome sizes, the number of unique genes and number of new genes per added new genome were calculated. Graphs showing the relationship between number of genomes and number of pan and core genome sizes, number of unique genes and number of new genes were constructed in R.

Determining whether a collection of genome sequences has a converging pan genome helps to understand the degree of plasticity and hence the genetic diversity of the genomes under consideration. To determine the convergence (or not) of pan genomes several methods, such as exponential and power regression equations have been proposed to model the pan genome expansion in relation to the number of genomes (Tettelin et al., 2008; Zhang and Sievert, 2014). In this thesis, the power law-regression equation was used to model the expansion of the both the *E. coli* and *K. pneumoniae* pan genomes as follows:

$$N = \beta g^{\gamma} \quad \text{(i)}$$

Where $N = number\ of\ genes\ in\ the\ pan - genome$

$g = number\ of\ pan - genome\ sequences$

β and γ are constants.

In order to determine the values of β and γ, equation (i) was transformed by taking the natural logarithms of both sides i.e.

$$\ln(N) = \ln(\beta g^{\gamma}) \quad \text{(ii)}$$

$$\Rightarrow \ln(N) = \ln(\beta) + \gamma \ln(g) \text{ (iii)}$$

Letting $\alpha = \ln(\beta)$, equation (iii) becomes:

$$\ln(N) = \alpha + \gamma \ln(g), \text{ (iv)}$$

The mean pan genome size of each set of 10 samples of g genomes (g=1,2,3…,G (G=overall total number of genomes), was then modelled against the pan genome size using the linear regression equation (iv) and the coefficients obtained to estimate α and γ. β was then calculated as $\beta = e^{\alpha}$

γ was the determinant coefficient such that:

if $\gamma > 0$, then equation (i) is unbounded and the pan genome would continue to expand (without converging) implying that addition of more genome sequences would bring new genes which had not been previously identified

if $\gamma \leq 0$, then equation (ii) is bounded and the pan genome would converge with some sufficient number of genomes implying, adding more genomes sequences of isolates from the same population would not result in identification of new genes.

All statistical analyses and plots were done using R v3.3.1 (R Core Team, 2014).

2.5 Mathematical modelling of force of *E. coli* BSI

With the overall aim of understanding the dynamics of drug resistant infection, mathematical modelling approach was undertaken to estimate the incidence of BSI, using the *E. coli* BSI as a case study.

To model the dynamics of CA-BSI, the community was considered as a black-box reservoir where a number of factors influence the dynamics of BSI and so many BSI cases remain undetected. However, it was assumed that the trends of CA-BSI detected in the hospital mimic trends in the community. It was hypothesised that for different reasons, people are born with different levels of susceptibility to infection. Two models of BSI dynamics, one assuming the population in Blantyre has heterogeneous susceptibility and another assuming the population as homogenous susceptibility to BSI, were developed and their fit to the observed age stratified incidence of *E. coli* BSI compared. The best fitting model was then used to model the incidence of E. coli BSI over time and generate as estimate for the force of infection.

Chapter Three: Trends in community acquired bloodstream infection and antimicrobial resistance in Blantyre

3.1. Overview

BSI is a common cause of morbidity and mortality in SSA yet few facilities are able to conduct long-term surveillance. MLW has conducted sentinel surveillance of bacteraemia at QECH since 1998 as described in Chapter Two. This chapter reports the long-term trends in BSI and AMR in Blantyre.

Analysis of the longitudinal blood culture surveillance data showed a significant decline in both prevalence and incidence of BSI infection in Blantyre especially after 2006. NTS and *S. pneumoniae* were the overall leading causes of BSI in Blantyre but their dominance diminished with time and were eventually overtaken by *S.* Typhi as the leading cause of BSI in 2012. The decline in BSI however, coincided with increasing rates of pathogens that were resistant to all the Malawian first-line antimicrobial agents especially among Gram-negative pathogens. Furthermore incidence of non-*Salmonellae* Enterobacteriaceae that were ESBL producing or fluoroquinolone resistant rose significantly. In contrast, most common Gram-positive pathogens remain susceptible to either penicillin or chloramphenicol, although emergence of MRSA was detected.

In this chapter, the term "other Enterobacteriaceae" has been used to mean all "Enterobacteriaceae species other than *Salmonellae*, *E. coli* and *K. pneumoniae".*

3.2. Introduction

Blantyre is a setting with high prevalence of HIV, malaria and malnutrition. However, for the past few years several interventions have been undertaken to reduce the burden of these conditions. Free antiretroviral therapy (ART) programmes as well as prevention of mother to child transmission (PMCT) programmes were initiated in June 2004 and 2005 respectively (Nyasulu, 2011; Makombe et al., 2007). Malaria control interventions such as distribution of free insecticide treated mosquito nets (Mathanga et al., 2012) have also been rolled out. Food security improved following the introduction of subsidised farm input programme in 2005 and community management of malnutrition (Denning et al., 2009; National Statistical Office (NSO) and ICF Macro, 2011). *Haemophilus* i*nfluenzae* type b vaccine and pneumococcal conjugate vaccine variant 13 (PCV13) were also introduced in 2002 and 2011 respectively (Bar-Zeev et al., 2015; Daza et al., 2006)

NTS, *S.* Typhi and *S. pneumoniae* were previously identified as leading causes of BSI in Blantyre (Gordon et al., 2008; Feasey et al., 2015c; Everett et al., 2011). Investigations into the epidemics of both NTS and *S*. Typhi revealed an association with multidrug resistance to the first-line antimicrobial agents (Gordon et al., 2008; Feasey et al., 2015c). The widespread MDR NTS necessitated the introduction and

increasingly extensive use of ciprofloxacin (from 2002) and ceftriaxone (from 2004), for the empirical management of sepsis (Gordon et al., 2008; Molyneux et al., 2009). Similar to trends in other parts of the world where cephalosporins and fluoroquinolones have been in use (Chimalizeni et al., 2010; Brink et al., 2012a; Kruger et al., 2004; Govinden et al., 2006; Holt et al., 2015), ESBL-producing and fluoroquinolone resistant *E. coli*, *Klebsiella and Salmonellae* isolates were detected in Blantyre following the extensive use of ceftriaxone and ciprofloxacin (Gray, 2006; Feasey et al., 2014b; Feasey et al., 2014a). However, the full burden of ESBL and fluoroquinolone resistance among the Gram-negative pathogens in this setting has yet to be described.

In contrast, fluctuating but relatively lower levels of resistance to penicillin ranging 9%-18% were previously reported in Gram-positive pathogens (Everett et al., 2011). Like in most SSA settings, prevalence of MRSA in Malawi remains unknown but it has recently been increasingly detected in Blantyre.

As previously described in this thesis (Section 1.8), surveillance has an important role in detecting the emergence of drug resistant organisms and tracking their spread. Long-term bacteraemia or AMR surveillance is however, rarely conducted in most SSA settings. Using the MLW sentinel surveillance dataset in this chapter, longitudinal trends in BSI and AMR over a 19-year period QECH have been described, with particular focus on prevalence of ESBL and fluoroquinolone resistance among Gram-negative pathogens and the emergence of MRSA.

3.3. Trends in Community acquired bloodstream I in Blantyre

3.3.1. General overview of blood cultures from QECH, 1998-2016

A total of 194,539 blood cultures were collected from patients presenting to QECH between January 1998 and December 2016. 115,44 (59.3%) cultures were taken from children where as 79,095 (41.0%) were taken from adult patients (Age distribution and age stratified incidence rates are presented in in section 3.4.4 of this chapter). 29,183 (15.0%) of cultures yielded significant pathogens, a further 36,763 (18.9%) yielded contaminants. No significant growth was detected from 128,593 (68.1%) blood cultures. Tables summarising number of isolates for each sub-group defined in Appendix 1 is presented in Appendix 2A-B. The absolute number of BCs collected per year fluctuated during the surveillance (Figure 1A) whilst number of confirmed BSI cases significantly declined (Figure 3.1).

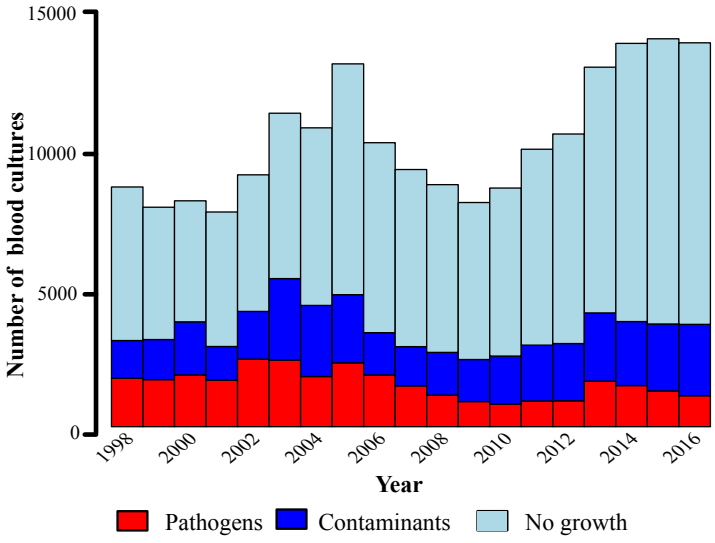

Figure 3.1 Distribution of surveillance blood cultures at QECH from 1998-2016. The number of blood cultures per year sub-categorised into significant pathogens (red), contaminants (blue) and no growth (light blue).

3.3.2. Causes of BSI in Blantyre

3.3.2.1. Enterobacteriaceae

Concurring with previous studies, which described epidemics of *Salmonellae* BSI in Blantyre (Feasey et al., 2015c; Gordon et al., 2008), NTS was the overall leading cause of BSI accounting for 10,473 (35.9%;) of the confirmed BSI cases (Table 2.1) and *S.* Typhimurium was the dominant serotype (Figure 3.1A) with 8,293/10,473 (79.2% NTS) while 1,658/10,178 (15.8%) NTS isolates were *S.* Enteritidis. The highest peak of NTS BSI was observed during the period 2002-2005 (Table 3.1) but had declined substantially especially after 2006 (Figure 3.2A) such that during the period 2010-2016, *S.* Typhi which had been rare before 2010, significantly rose to become the leading cause of BSI in this setting (Figure 3.2A).

There were a total of 2,134 *S*. Typhi-confirmed blood cultures (8.0% BSI) during the surveillance but the majority 1,948/2,134 (91.3%) were isolated between 2010 and 2016 (Table 3.1). The peak of the S. Typhi epidemic was in 2013, where 887 cases were recorded but has since declined to 298 in 2016 (Figure 3.2A).

A total of 2,5600 (8.8% BSI) *E. coli* were isolated during the surveillance period, making *E. coli* the overall third most common causes of BSI in Blantyre (Table 3.1). The highest number of *E. coli* BSI cases recorded in a calendar year was 193, in 2005 (Figure 3.2B). There was an average of 135 *E. coli* BSI cases per year with no evidence of epidemics during the study period but in each year *E. coli* was consistently among the top four causes of BSI in Blantyre (Figure 3.2B).

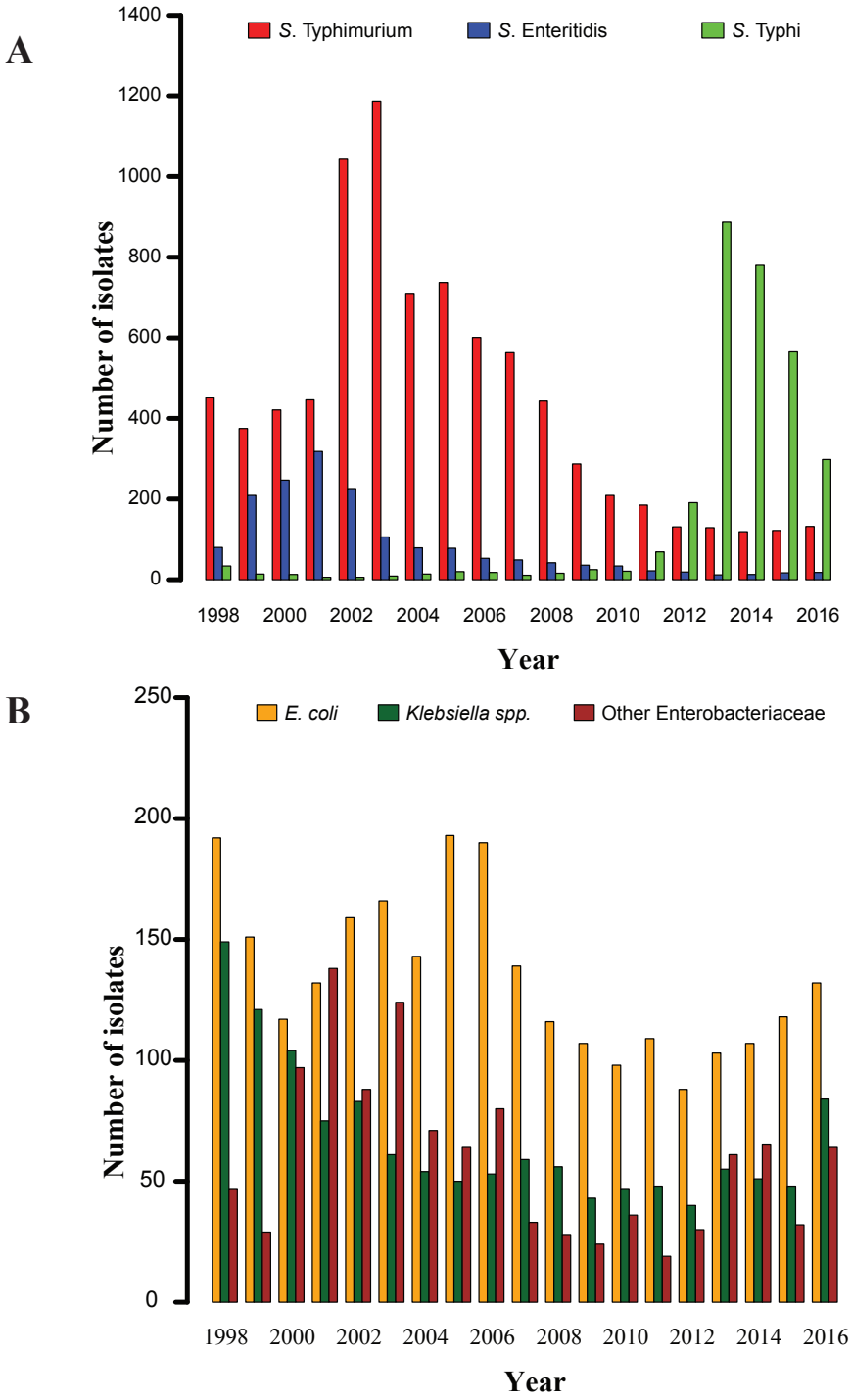

Figure 3.2 Distribution of Enterobacteriaceae isolates from 1998-2016: (A) *Salmonellae* (B) Non-*Salmonellae* Enterobacteriaceae.

There were 1,1281 isolates of *Klebsiella spp.*, representing 4.4% of all pathogens isolated during the study period. The highest number of cases (149) was recorded in 1998. Between 1998 and 2000, over 100 cases were recorded each year but had since fallen to an average of 57 cases per year (Figure 3.2B).

Other Enterobacteriaceae species were responsible for 1,130 culture-confirmed BSI cases (3.9% BSI). The most common organisms amongst the other members of the Enterobacteriaceae were *Enterobacter spp*., 517 (45.8%), *Serratia spp*., 211 (18.7%), *Citrobacter spp*. 185 (16.4%), *Proteus spp*., 93, (8.2%) and *Shigella spp*., 40 (3.5%) (Table 3.1).

Table 3.1 Prevalence of significant pathogens in four time intervals: 1998-2001, 2002-2005, 2006-2009 and 2010-2013, 2014-2016

Organism	**No. of isolates (%)**					
	1998-2001	2002-2005	2006-2009	2010-2013	2014-2016	Total
Acinetobacter spp	200 (3.0)	126 (1.0)	115 (2.0)	64 (1.5)	40 (1.0)	545 (1.9)
Anaerobes	7 (0.0)	8 (0.0)	7 (0.0)	6 (0.1)	2 (0.1)	30 (0.1)
Citrobacter spp	69 (1.0)	75 (0.8)	20 (0.4)	9(0.2)	12 (0.3)	185 (0.6)
E. coli	592 (9.0)	661 (7.0)	552 (10.0)	398 (9.3)	357(9.3)	2560 (8.8)
E. faecalis	48 (0.7)	57 (0.6)	61 (1.1)	27 (0.6)	27(0.7)	220 (0.8)
Edwardsiella	-	2 (0.0)	-	-	-	2 (0.0)
Enterobacters pp	93 (1.3)	173 (2.0)	88 (1.6)	74 (1.7)	89(2.3)	517 (1.8)
Enterococcus spp	14 (0.2)	2 (0.0)	2 (0.0)	64 (1.5)	69 (1.8)	151 (0.5)
Escherichia spp	-	3 (0.0)	3 (0.1)	-	-	6 (0.0)
Flavobacteria spp.	2 (0.0)	2 (0.0)	2 (0.0)	-	-	6 (0.0)
H. influenza type b	112 (1.6)	92 (1.0)	26 (0.5)	30 (0.7)	15 (0.4)	275 (0.9)
Haemophilus spp.	41 (0.6)	53 (0.6)	36 (0.7)	21 (0.5)	8 (0.2)	434 (0.5)
Hafnia spp.	3 (0.0)	2 (0.0)	1 (0.0)	-		6 (0.0)
Klebsiella spp.	449 (7.0)	248 (3.0)	211 (4.0)	190 (4.4)	183 (4.8)	1281 (4.4)
Kluyvera spp.	2 (0.0)	1 (0.0)	2 (0.0)	2 (0.0)	-	7 (0.0)
M. morganii	-	3 (0.0)	4 (0.1)	11 (0.3)	1 (0.0)	19 (0.1)
Mycobacterium	60 (1.0)	-	-	-		60 (0.2)
Neisseria	78 (1.0)	27 (0.0)	42 (1.0)	34 (0.8)	39(1.0)	220 (0.8)
Other Gram-negative	33 (0.5)	56 (1.0)	52 (1.0)	56 (1.3)	7(0.2)	204 (0.7)
NTS	2,685 (38.9)	4,432 (50.1)	2,141 (40.2)	782 (18.3)	433 (11.3)	10473 (35.9)
Streptococcus spp.	115 (1.7)	113 (1.3)	55 (1.0)	46 (1.1)	57 (1.5)	386 (1.3)
Pantoea spp.	-	-	-	11(0.3)	9 (0.2)	20 (0.1)
Proteus spp.	47 (0.7)	9 (0.1)	12 (0.2)	7 (0.2)	18 (0.5)	93 (0.3)
Pseudomonas	98 (1.0)	102 (1.0)	41 (1.0)	126 (2.9)	75 (2.0)	442 (1.5)
Raoultella spp.	-	-	-	3 (0.1)	8 (0.2)	11 (0.0)
S. agalactiae	173 (2.5)	155 (1.8)	40 (0.8)	70 (1.3)	17 (0.4)	455 (1.6)
S. aureus	505 (7.0)	480 (5.0)	258 (5.0)	344 (8.0)	338 (8.8)	1925 (6.6)
S. pneumoniae	1,139 (17.0)	1,476 (17.0)	1,072 (20.0)	448 (10.5)	123 (3.2)	4258 (14.6)
GAS	117 (1.7)	102 (1.2)	69 (1.3)	59(1.4)	50 (1.3)	397 (1.4)
S. Typhi	67 (1.0)	49 (1.0)	70 (1.0)	1168 (27.3)	1643 (43.0)	2997(10.3)
Serratia spp.	92 (1.3)	66 (0.7)	15 (0.3)	21 (0.5)	17 (0.4)	211(0.7)
Shigella spp.	1 (0.0)	9 (0.1)	18 (0.3)	6 (0.1)	6 (0.2)	40 (0.1)
Vibrio spp.	5 (0.0)	2 (0.0)	2 (0.0)	3 (0.1)	-	12 (0.0)
Yersinia	2 (0.0)	2 (0.0)	1 (0.0)	2 (0.0)	-	7 (0.0)
Candida spp.	4 (0.0)	6 (0.1)	4 (0.1)	13 (0.3)	13 (0.3)	40 (0.1)
Cryptococcus spp.	50 (1.0)	249 (2.8)	300 (5.6)	195 (4.6)	169 (4.4)	963 (3.3)
Total	**6,903**	**8,843**	**5,322**	**4290**	**3825**	**29183**

3.3.2.2. Other Gram-negative pathogens

Other Gram-negative causes of BSI included *Acinetobacter spp.* (545), *Haemophilus* spp. (434), *Pseudomonas* spp. (442), *Neisseria* spp. (220), other Anaerobes (30) and *Vibrio* spp. (12). These together were responsible for 1,553 (5.7%) positive blood cultures (Table 3.1). Annual number of BSI cases by pathogens in this group declined during the surveillance mostly due to decline in cases of *H. influenzae* and *Acinetobacter* spp. (Figure 3.3). Substantial declines in *H. influenzae* were observed after 2002, when the *H. influenzae* type b conjugate vaccine was introduced (Daza et al., 2006).

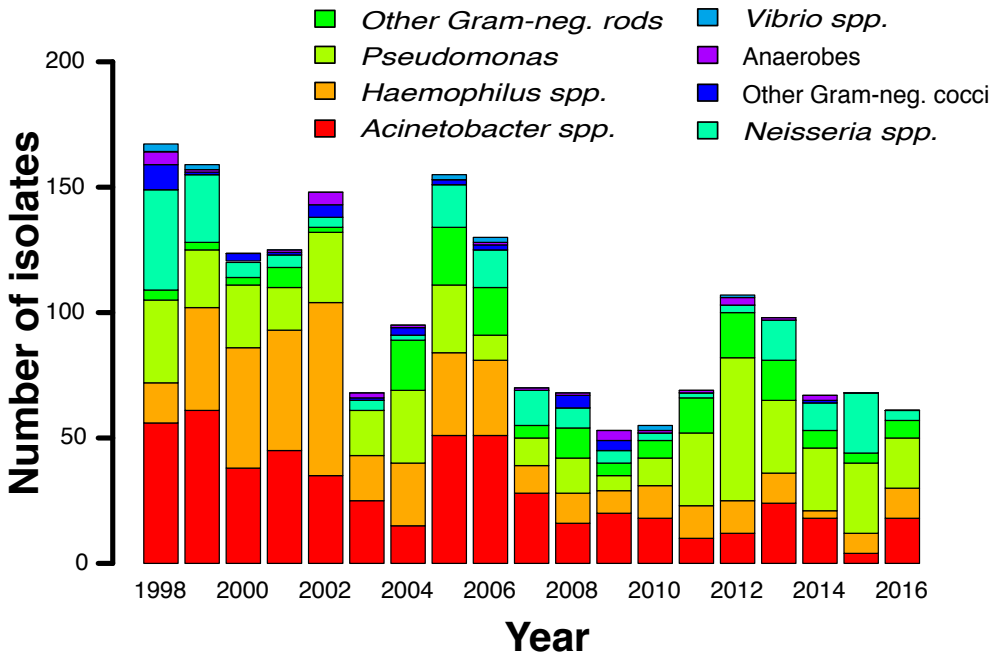

Figure 3.3 Distribution of BSI cases by non-Enterobacteriaceae Gram negative pathogens in Blantyre, 1998-2016.

3.3.2.3. S. pneumoniae

Trends in prevalence of *S. pneumoniae* BSI were previously described for the period 2000-2009 (Everett et al., 2011), whereas this dataset spans 1998-2016.

S. pneumoniae were the second most common cause of BSI in Blantyre with 4,258 *S. pneumoniae* (14.6% BSI) isolated between 1998 and 2016. The highest frequency of *S. pneumoniae* was in the period 2002-2005 but similar to NTS, there were substantial declines in number of *S. pneumoniae* BSI from 2006 (Figure 3.4A).

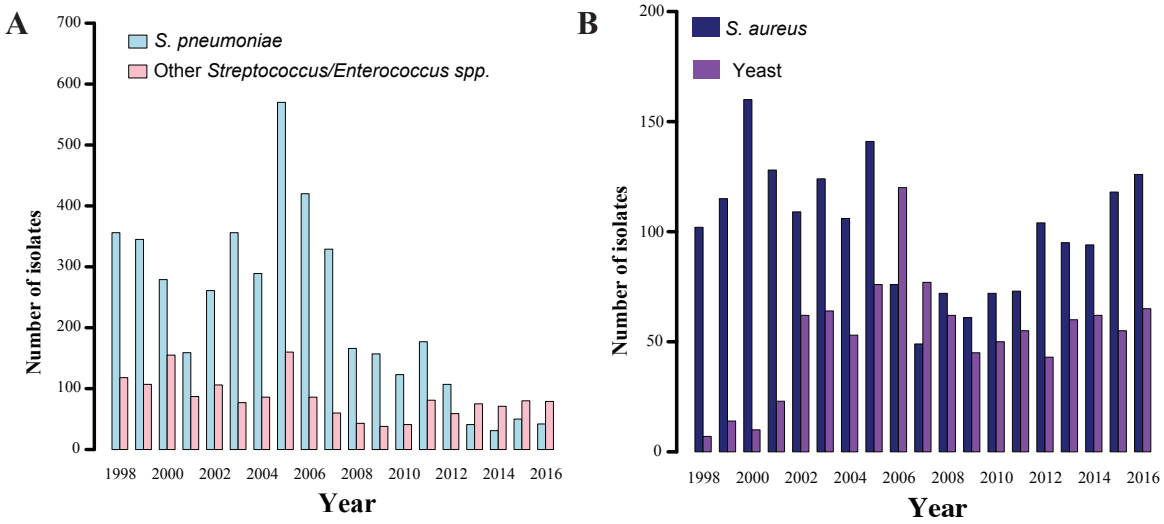

Figure 3.4 Distribution of BSI cases by Gram positive pathogens and yeast in Blantyre, 1998-2016. (A) Shows distribution of BSI by *S. pneumoniae* and (B) BSI by other *Streptococcus/Enterococcus spp*.

3.3.2.4. S. aureus

S. aureus was the second most common Gram-positive cause of BSI with 1,923 isolates (6.6% BSI) and the isolation frequency of fluctuated throughout the study period (Figure 3.4B).

3.3.2.5. Other Gram-positive pathogens

There were a total of 1,609 beta-haemolytic *Streptococci and Enterococci* (5.4% BSI) isolate, constituting 1,238 (4.5% BSI) β-haemolytic *Streptococcus* spp*.* and 371 *Enterococci.* Among the β-haemolytic Streptococci there were 397 (1.4% BSI) isolates of Lancefield Group A, 455 (1.6% BSI) isolates of Group B *Streptococci* and 386 other *Streptococci.* On the other hand, the 371 *Enterococci* comprised of 220 *E. faecalis,* 76 *E. faecium* and 75 other *Enterococcus* spp. (Table 3.1).

3.3.2.6. Yeasts

The longitudinal bacteraemia surveillance also led to the identification of 883 yeast isolates (3.4% BSI; Table 3.1), including 849 *Cryptococcus neoformans* (96.5%) and 30 *Candida* spp*.* (3.4%). Similar to *S. aureus*, annual BSI cases due to yeast fluctuated during the surveillance (Figure 3.4).

3.3.3. Trends in incidence of BSI

Minimum incidence rates for each of the high frequency (pathogens with ≥ isolates in at least any one year) and medium frequency pathogens (pathogens with 50-299 isolates per years) were estimated as previously described in Chapter Two and shown in Figure 3.2 below. Due to low numbers of BSI cases per year, minimum

incidence rates were not estimated for other Gram-negative pathogens with low frequency.

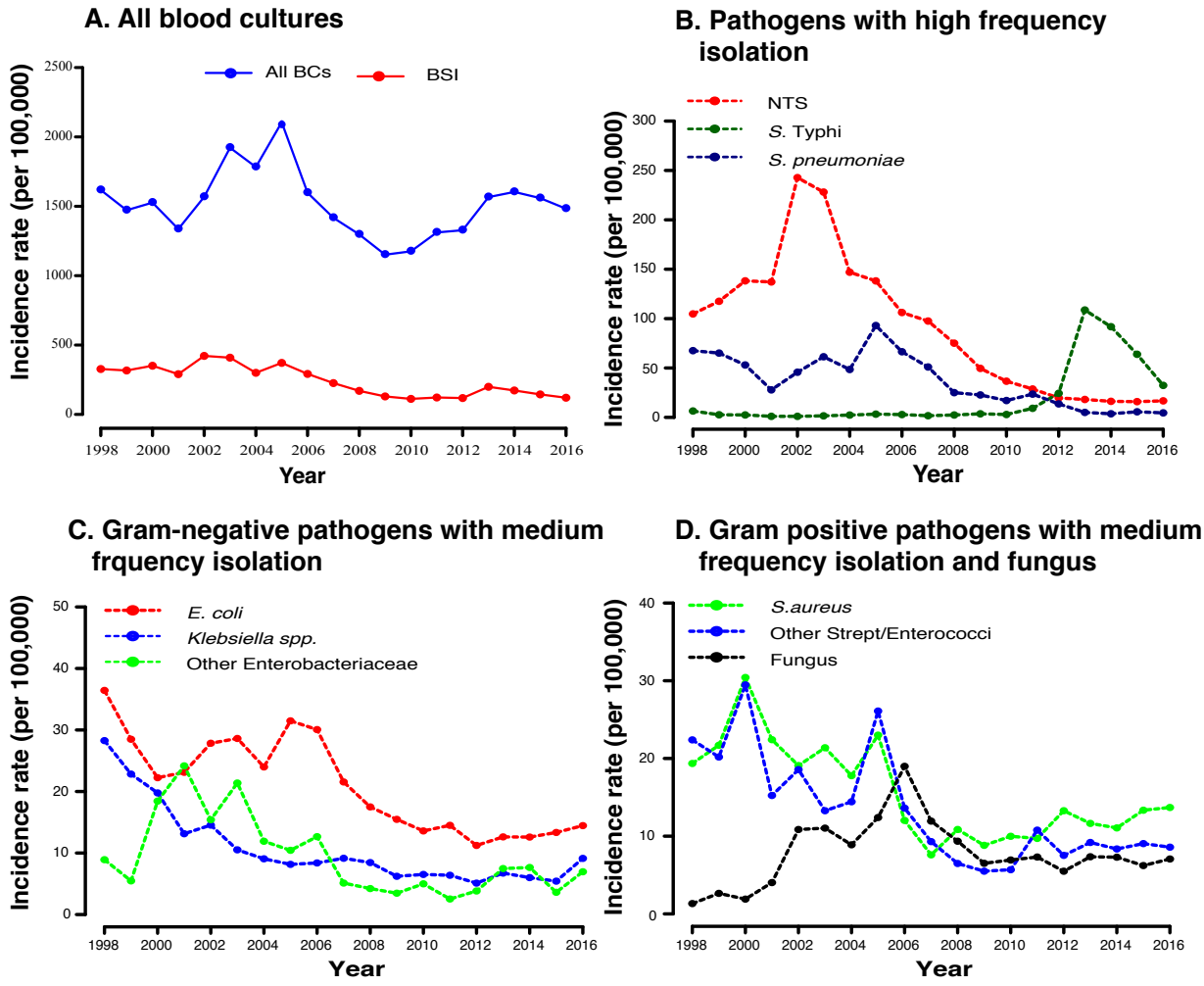

Figure 3.5 Trends in incidence of BSI 1998-2016. (A) Annual frequency of blood culture sampling and pathogen (B) Estimated minimum incidence of pathogens isolated at high frequency (≥300/year). (C) Estimated minimum incidence of Gram-negative pathogens isolated with medium frequency (50-299/year) (D) Estimated minimum incidence of Gram-positive and yeast pathogens isolated with medium frequency.

Although a significant increase in absolute numbers of blood cultures per year was observed, there was a significant declining trend in proportion of blood cultures to population (p<0.0001). A negative binomial regression model of the estimated minimum incidence of BSI showed declining trend in overall incidence of BSI (p<0.0001) during the study period and the decline was markedly observed after from 2006 (Figure 3.5A). This declining trend was also evident in all the high and medium frequency pathogens (all p<0.0001) except for *S.* Typhi, which until 2010 had a roughly uniform incidence (~0; Figure 3.5B). It is also important to note that despite a general decline, incidence of *E. coli* BSI increased from 2012 to 2016 (Figure 3.5C).

3.3.4. Age distribution of patients with confirmed CA-BSI

Of the 29,183 episodes of culture-confirmed BSI, age was recorded in 23,219 (79.6%). 13,002 (56.0%) cases with known age were children (<16 years) and 10,217 (42.5%) were adults (≥16 years old). 10,059 (77.4%) of children were aged below five years. Most bacterial species associated with BSI in Blantyre displayed a bimodal age distribution, affecting children under five years of age and adults aged between 20-45 years (Figure 3.6A-B&D-H). The only exception was *S*. Typhi, which was most common in children under ten years of age (Figure 3.6C). Cryptococcal BSI was most common in adults aged between 20 and 45 years, but uncommon in

children (Figure 3.6I).

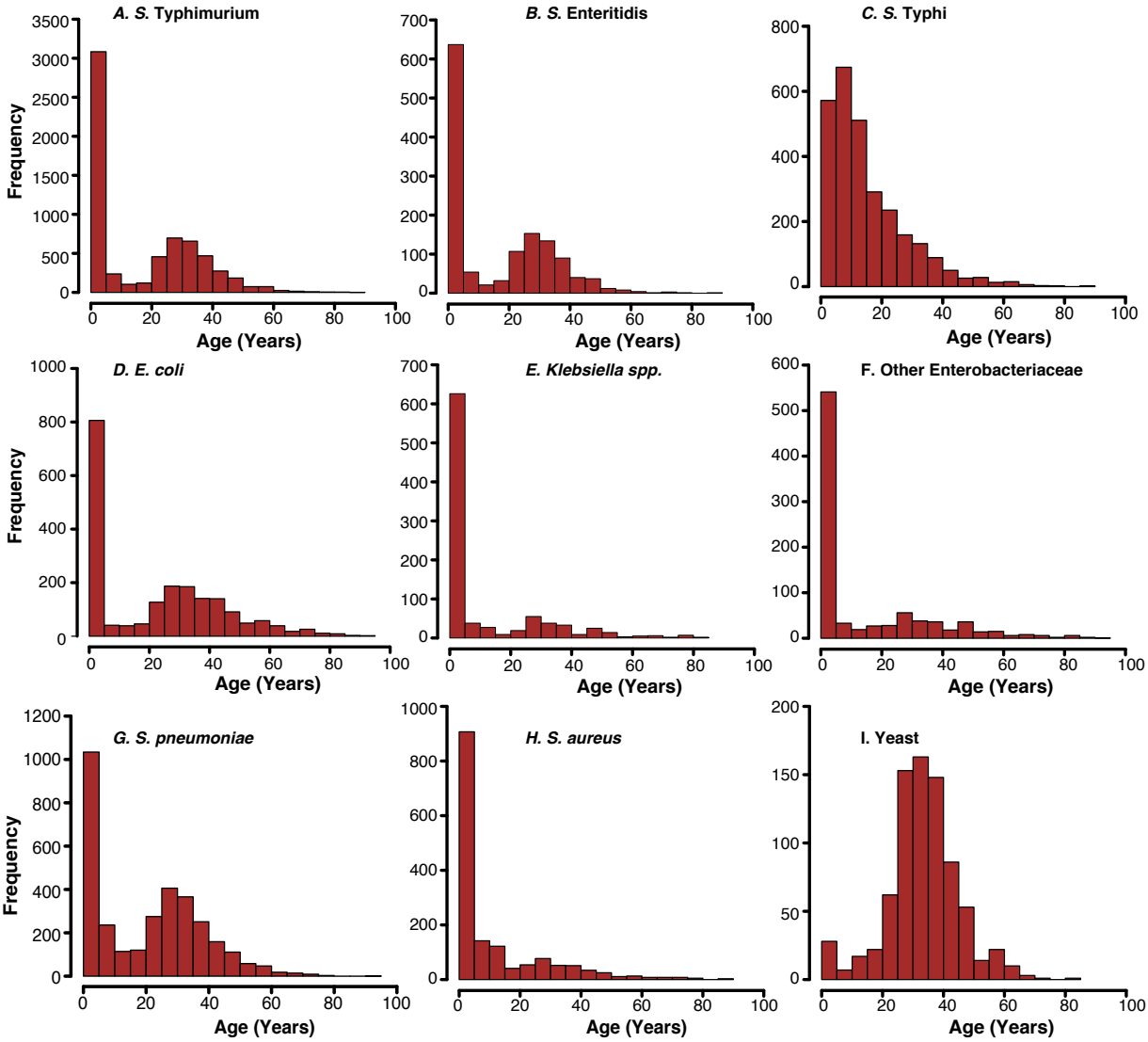

Figure 3.6 Distribution isolates by age of patients with BSI. Each graph represents specific pathogen isolates as follows: (A) *S.* Typhimurium, (B) *S.* Enteritidis, (C) *S.* Typhi, (D) *E. coli*, (E) *Klebsiella spp*. (F) Other Enterobacteriaceae, (G) *S. pneumoniae*, (H) *S. aureus* and (I) Yeast in Blantyre from 1998-2016.

3.3.4.1. Age stratified incidence

When the aggregate data were adjusted by age distribution in the population to produce minimum incidence estimates stratified by age, a subtly different picture from the absolute number of cases emerged. It was noted that the minimum incidence rates for *E. coli* BSI were greatest in older age groups ≥ 70 years (up to 54.3/100,000/year) followed by children aged below five years (37.5/100,000/year); and for *Klebsiella* BSI, incidence rates were highest in children below five years (29.2/100,000/year) followed by the elderly aged 75-80 years (24.6/100,000/year). For *S*. Typhi, incidence rates were greatest in children between 5-10 years and for all other bacterial pathogens they were greatest in those below 5 years of age (Figure 3.7).

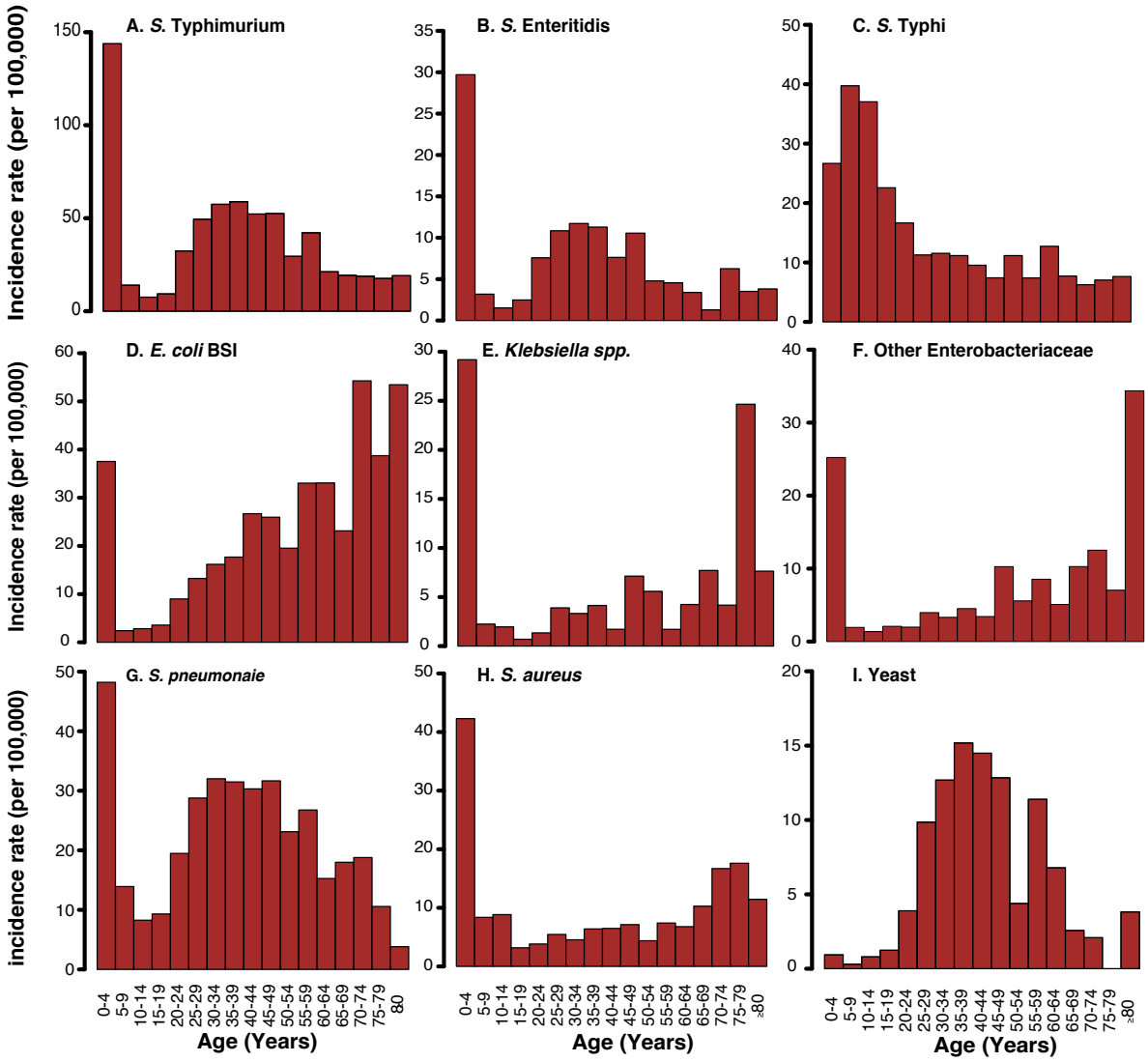

Figure 3.7 Estimated minimum incidence rates of BSI stratified by age for (A) *S.* Typhimurium (B) *S*. Enteritidis (C) *S*. Typhi, (D) *E. coli*, (E) *Klebsiella spp*., (F) other Enterobacteriaceae, (G) *S. pneumoniae*, (H) *S. aureus* (I) Yeast.

3.3.5. Contaminants

The following organisms were classified as contaminants: Coagulase negative *Staphylococcus*, *Micrococuss spp*., *Bacillus spp*., *Diptheroids*, alpha-haemolytic *Streptococcus* and Skin flora (Appendix 1).

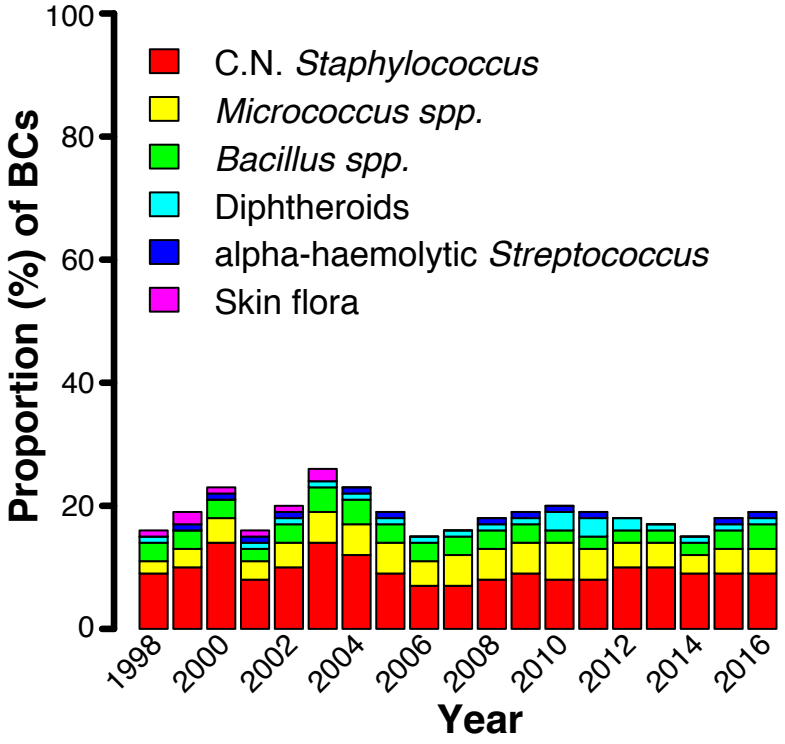

Figure 3.8 Distributions of contaminants by organism, expressed as proportions of all blood cultures per year of isolation.

Of the 36, 763 isolates identified as contaminants, 18,675 (50.8%) isolates were Coagulase negative *staphylococcus,* making them the most common contaminants in this setting. Other common organisms included *Micrococcus spp.* (8,187 [22.3%]), *Bacillus spp*. (5,631 [15.3%]) and Diptheroids [2,056 [5.6%]). The overall

contamination rate was 18.9% but this varied by year. The general trend of contamination rates was increasing (p<0.0001) but as it can be seen (Figure 3.8), the distribution of these rates over time was distinct between two periods: 1998-2005 and 2006-2016. Contamination rates were significantly higher (p<0.0001) and relatively more dispersed during 1998-2005 (range=15.7-26.0; mean rate=20.1) than contamination rates for 2006-2016 (range =14.8-20.1, mean rate=14.8). By age group, blood cultures from children had higher contamination rates than blood cultures from adults (95%CI=[0.046,0.053], p<0.0001).

3.4. Trends in Antimicrobial resistance

27,249/28,180 (96.7%) confirmed bacterial BSI isolates were tested for susceptibility to at least one antimicrobial agent and 25,752 (91.4%) BSI isolates were tested for susceptibility to at least three antimicrobial agents. 13,343/25,572 (52.2%) isolates tested for susceptibility to at least three first-line agents, including ampicillin (penicillin for Gram-positive pathogens), cotrimoxazole and chloramphenicol were RFL, 10,316/25,572 (40.3%) were resistant to one or two first-line agents, and 2,093 isolates (8.2%) were susceptible to all three first-line agents. Overall, proportions of RFL isolates followed an upward trend during the surveillance period. RFL was markedly more common among Gram-negative isolates (12,902/18,887 [68.3%]) than Gram-positive isolates (441/6,685 [6.6%]). 6,129/6,685 (91.7%) of all Gram-positive isolates were susceptible to either penicillin or chloramphenicol.

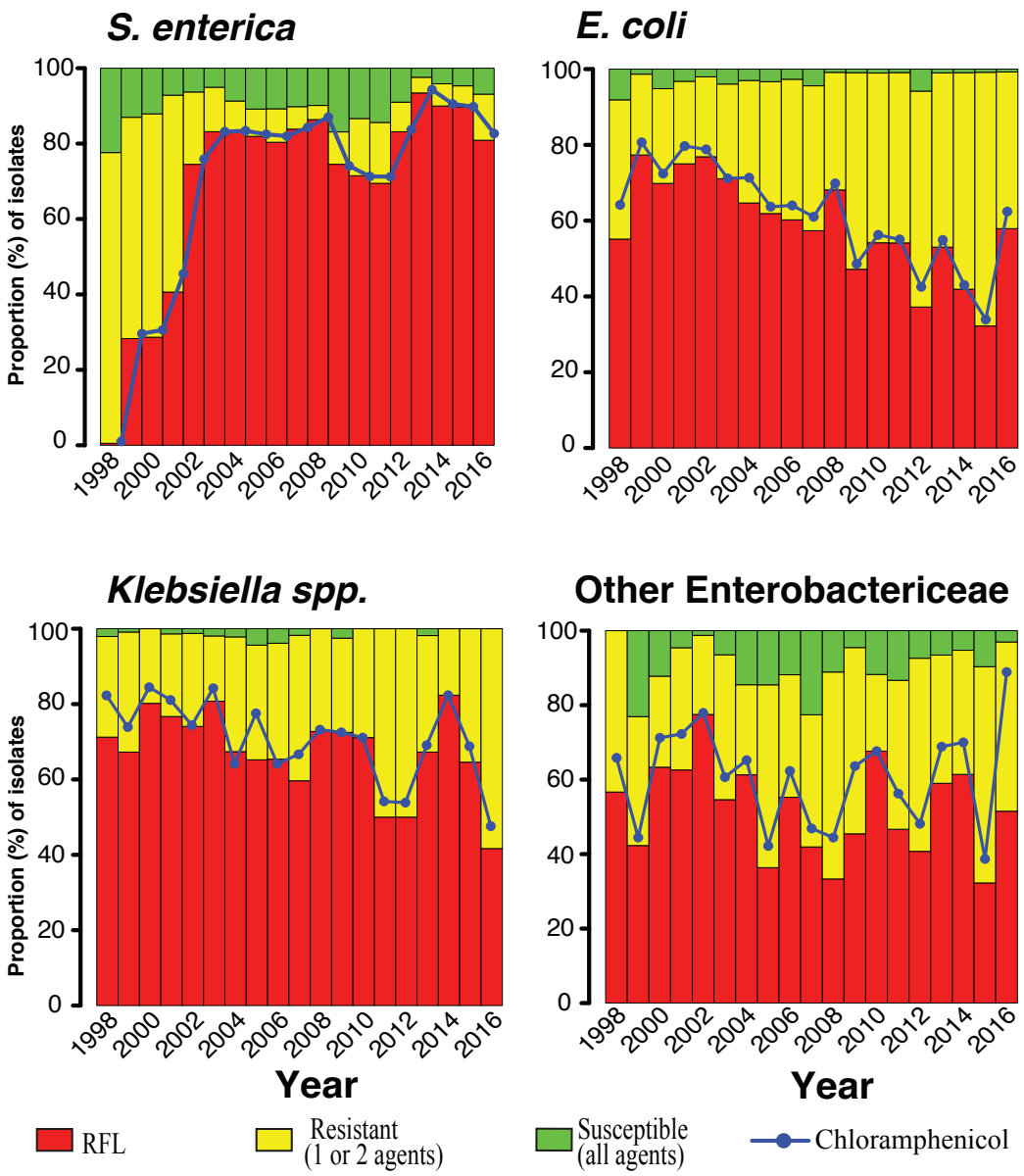

Figure 3.9 Trends in proportions of isolates resistant to Malawian first line antimicrobials (ampicillin, cotrimoxazole and chloramphenicol) in Enterobacteriaceae pathogens.

3.4.1. Partial return of chloramphenicol susceptibility in Enterobacteriaceae

In contrast to the overall trends in RFL isolates, proportions of RFL *E. coli* substantially declined during the study period (p <0.0001), primarily due to a decline in chloramphenicol resistance (Figure 3.9B; p<0.0001). A decline in chloramphenicol resistance was also observed in *Klebsiella spp.* (Figure 3.9C; p<0.0001) but increased in *Salmonella* isolates (Figure 3.9A; p<0.0001). No significant trend in proportion of RFL isolates was detected in other members of the Enterobacteriaceae family (Figure 3.9D; p=0.22).

3.4.2. Emergence of ESBL producing, fluoroquinolone and aminoglycoside resistant isolates

ESBL production was first detected in *E. coli* in 2004 (1/142 [0.7%], in *Klebsiella spp*. (2/17 [11.7%]) and other Enterobacteriaceae (7/27[25.9%]) in 2003. A Cochran-Armitage trend test for proportions showed significant increases in third generation cephalosporin resistance rates in all non-*Salmonellae* Enterobacteriaceae (Figure 3.10). In addition to the Enterobacteriaceae, a majority (168/274 [61.3%]) of *Acinetobacter spp.* isolates were also ESBL producing with an increasing trend in proportions of ESBL producing *Acinetobacter* spp. isolates observed over time. Ceftazidime was not widely available. Therefore, *Pseudomonas* isolates were not routinely tested. However, 7/9 *Pseudomonas* isolates tested were resistant to ceftazidime.

Ciprofloxacin resistance was first detected in Blantyre in *Acinetobacter spp*. (1/43 [2.3%]) isolates in 2001 and in *E. coli* (4/22 [2.5%)] and *Klebsiella* isolates in 2003. As with resistance to third generation cephalosporins, an increasing trend in proportions of non-*Salmonellae* Enterobacteriaceae isolates with resistance to ciprofloxacin was observed (Figure 3.10). Ciprofloxacin resistance was detected in 105/393 (26.7%) *Acinetobacter spp.* and in 55/344 (12.8%) *Pseudomonas* isolates over the entire surveillance period.

During the surveillance, 462/2536 (18.2%) *E. coli*, 565/1265 (51.9%) *Klebsiella spp.*, and 320/1076 (29.7%) other Enterobacteriaceae pathogens resistant to gentamicin were isolated with substantial increases in proportions of resistant isolates observed over time (Figure 3.8). In other Gram-negative pathogens, high proportions of *Acinetobacter spp.* 201/504 (39.9%) and *Pseudomonas spp.* 127/370(34.3%) isolates were also resistant to gentamicin.

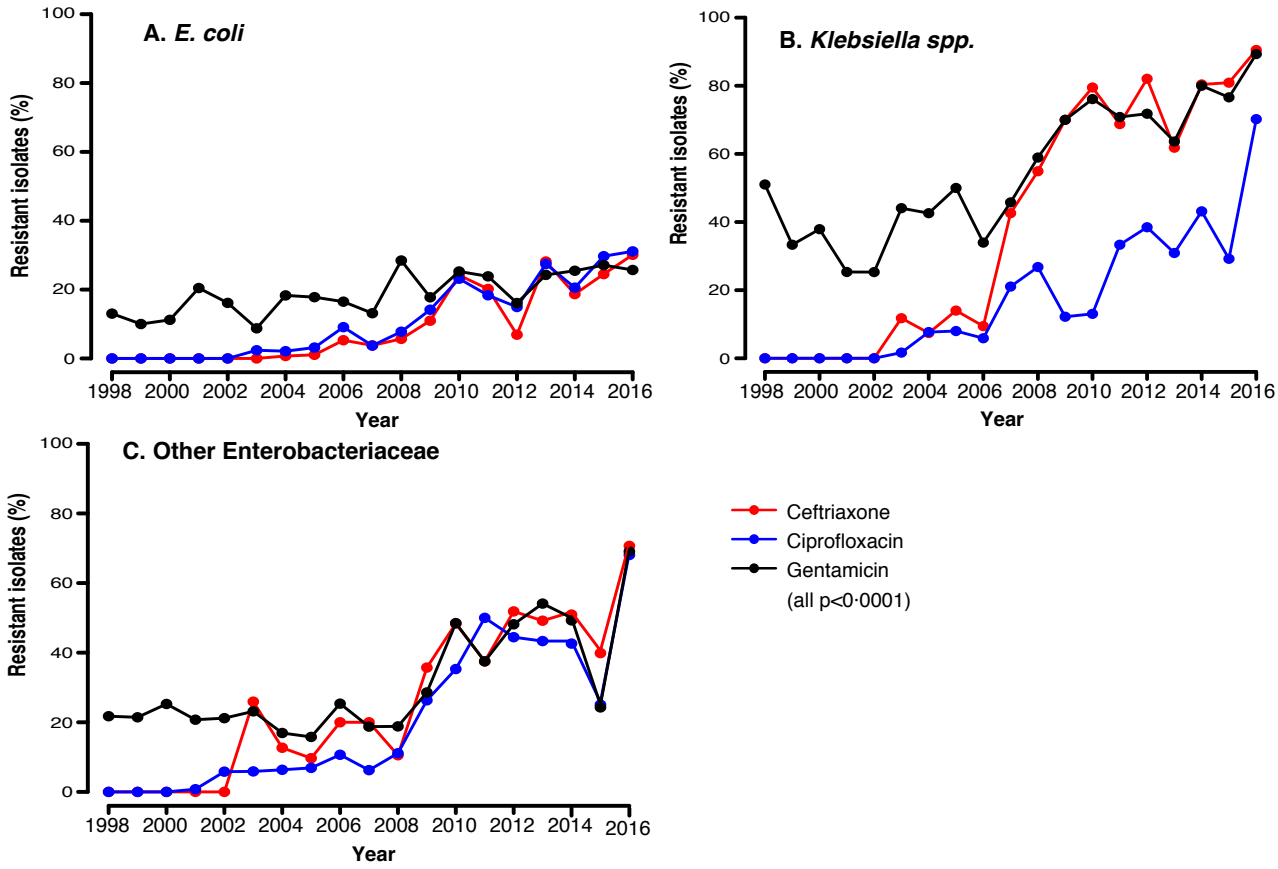

Figure 3.10 Trends in resistance to second-line antimicrobial agents in (A) *E. coli*, (B) *Klebsiella spp*. and (C) other Enterobacteriaceae

3.4.3. Antimicrobial resistance in Gram-positive pathogens

3.4.3.1. S. pneumoniae

Only 37/4,049 (0.9%) *S. pneumoniae* isolates were RFL. Resistance to cotrimoxazole (33,780/4,087 isolates [92.5%]) was almost universal. In contrast only 610/4,043 (15.1%) isolates were resistant or intermediate resistant to penicillin (falling to 551 [13.6%] when intermediate resistance was excluded). The overall trend in penicillin

resistant isolates was not significant (p=0.23) but a marked rise was observed following the introduction of the PCV13 in 2011 (Figure 3.11A; p<0.0001). 1,074/4,111 (26.1%) *S. pneumoniae* isolates were resistant to chloramphenicol, a further 2,200/4,080 (53.9%) *S. pneumoniae* isolates were resistant to tetracycline where as only 92/4,107 (2.2%) *S. pneumoniae* isolates were resistant to the macrolide erythromycin.

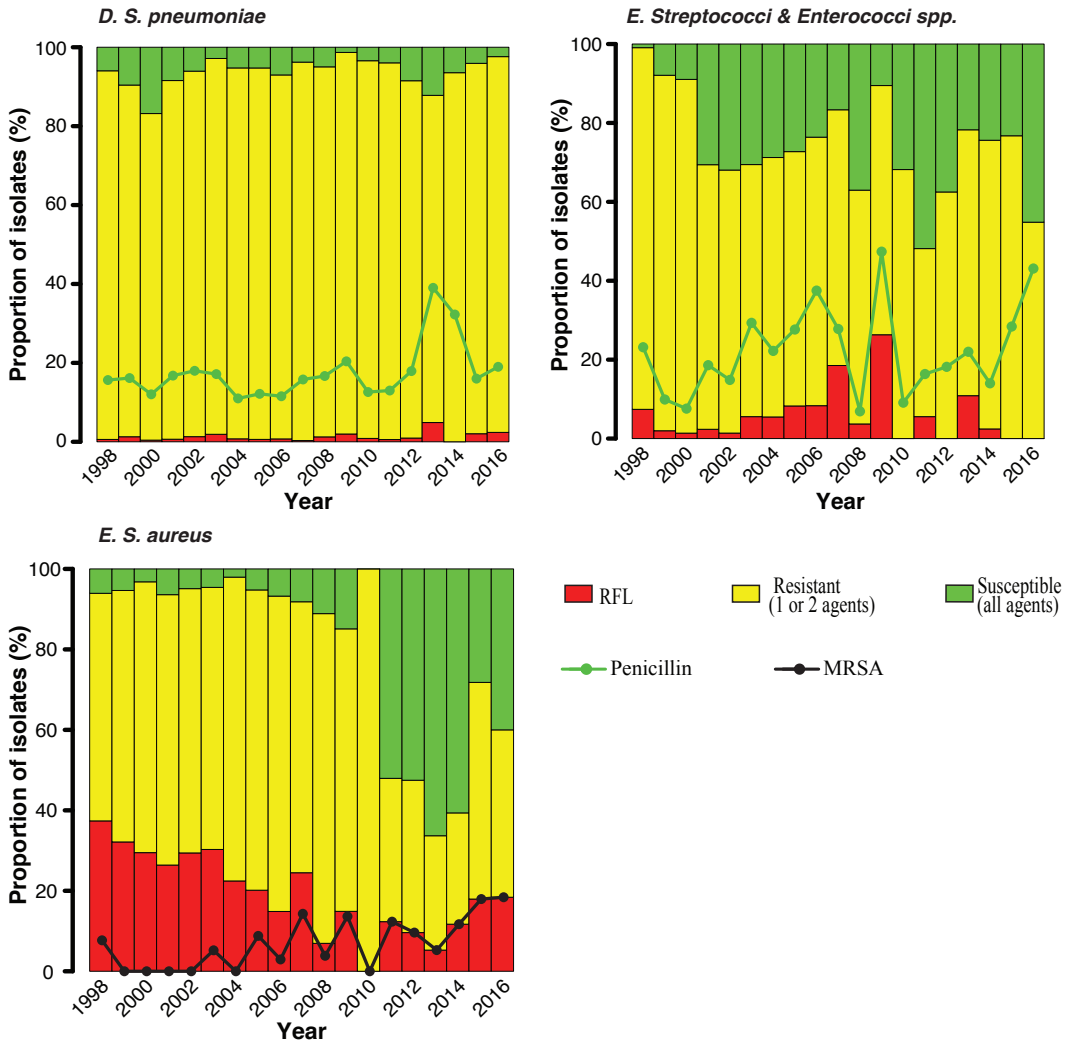

Figure 3.11 Trends in resistance to first-line agents (penicillin, cotrimoxazole and chloramphenicol) in Gram-positive pathogens.

3.4.3.2. MRSA

In total, 1,105/1,215 (79.0%) *S. aureus* isolates were resistant to penicillin whilst 828/1,895(43.7%) were resistant to cotrimoxazole and 452/1,898 (26.7%) to chloramphenicol. 107/1,118(9.6%) *S. aureus* isolates tested were MRSA. MRSA was first detected in 1998, but was not regularly isolated until 2005 (Figure 3.11C). Only 20/1,681 (1.2%) *S. aureus* isolates were tested for susceptibility to ciprofloxacin and none were resistant. 206/1,790 (10.6%) isolates tested were resistant to gentamicin.

3.4.3.3. Other Streptococcus and Enterococcus spp.

855/1,213 (70.5%) of the β-haemolytic *Streptococci* isolates and 229/325 (70.5%) *Enterococci* isolates were resistant to cotrimoxazole, whereas 176/1149 (15.3%) and 108/176 (61.4%) *Streptococci* and *Enterococci* spp. were resistant to penicillin and ampicillin respectively 208/1143 (17.1%) *Streptococci* and 198/326 (60.7%) *Enterococci* were resistant to chloramphenicol. Amongst Group A *Streptococci*, 14/382 (3.7%) were penicillin resistant, however these isolates were not speciated and therefore *Streptococcus pyogenes* isolates could not be specifically identified within this group.

3.4.4. Multi-drug resistance

The majority of Gram-negative pathogens including *E. coli*, *Klebsiella* spp., other non-*Salmonellae* Enterobacteriaceae, *Haemophilus* spp. and *Acinetobacter* spp. were found to be MDR. Relatively less but substantial proportions of *S. pneumoniae*, *S. aureus* and other *Streptococcus* spp. and *Enterococcus spp.* were also MDR.

Proportions of MDR isolates in *Klebsiella spp., Streptococcus spp.* and *Enterococcus spp., H. influenzae and Acinetobacter spp.* had an increasing trend (p<0.0001). However, declining MDR trend was detected in *E. coli* (p=0.001), whereas no significant trends were observed in other Enterobacteriaceae, *S. pneumoniae* and *S. aureus* (Table 3.2A & B). Almost all *Salmonellae* isolates were susceptible to ceftriaxone and ciprofloxacin, hence MDR phenotype in *Salmonellae* was determined by their RFL phenotype, i.e. MDR rates in *Salmonellae* were equivalent to the RFL rates depicted in Figure 3.9A.

Table 3.2A Frequency of multidrug resistance (MDR) in Gram-negative pathogens causing BSI in Blantyre between 1998 and 2016.

Year	*E. coli*	*Klebsiella spp.*	Enterobacteriaceae	*Acinetobacter*	*Haemophilus spp.*
1998	111/185(60)	66/146 (45.2)	22/30 (73.3)	26/53 (49.1)	1/16 (6.3)
1999	118/150(78.7)	31/113 (27.4)	12/26 (46.2)	36/61 (59.0)	12/28 (42.9)
2000	83/116 (71.6)	38/101 (37.6)	60/90 (66.7)	19/35 (54.3)	11/31 (35.5)
2001	96/124 (77.4)	17/73 (23.3)	87/131 (66.4)	23/43 (53.5)	26/47 (55.3)
2002	113/147 (76.9)	20/81 (24.7)	64/80 (80.0)	14/33 (42.4)	26/49 (53.1)
2003	110/152 (72.4)	24/52 (46.2)	67/108 (62.0)	14/23 (60.9)	11/14 (78.6)
2004	92/133 (69.2)	21/46 (45.7)	39/62 (62.9)	8/13 (61.5)	14/20 (70.0)
2005	121/181 (66.9)	22/46 (47.8)	24/55 (43.6)	26/45 (57.8)	10/23 (43.5)
2006	84/186 (45.2)	18/52 (34.6)	45/76 (59.2)	41/50 (82.0)	7/24 (29.2)
2007	81/136 (59.6)	27/57 (47.4)	15/31 (48.4)	20/27 (74.1)	6/10 (60.0)
2008	81/116 (69.8)	36/55 (65.5)	10/27 (37.0)	12/16 (75.0)	6/12 (50.0)
2009	60/106 (56.6)	29/40 (72.5)	12/22 (54.5)	11/18 (61.1)	9/9 (100.0)
2010	64/96 (66.7)	34/45 (75.6)	25/34 (73.5)	11/15 (73.3)	7/12 (58.3)
2011	76/109 (69.7)	35/48 (72.9)	10/15 (66.7)	7/9(77.8)	6/13 (46.2)
2012	42/86 (48.8)	32/38 (84.2)	17/27 (63.0)	11/12 (91.7)	11/13 (84.6)
2013	74/100 (74.0)	33/55 (60.0)	42/61 (68.9)	17/24 (70.8)	7/12 (58.3)
2014	63/105 (60.0)	41/51 (80.4)	40/57 (70.2)	14/17 (82.4)	1/3 (33.3)
2015	65/118 (55.1)	39/48 (81.3)	10/33 (78.5)	3/3 (100)	6/8 (75.0)
2016	92/133 (69.2)	77/84 (91.7)	51/65 (78.5)	13/16 (81.3)	7/12 (58.3)
Overall	**1626/2479 (65.9)**	**640/1231(47.7)**	**652/1030 (63.4)**	**310/494 (62.8)**	**171/336 (51.0%)**
***p*-value**	**0.001***	**<0.001****	**0.747**	**0.0044****	**<0.001****

*=Decreasing trend; **=increasing trend

Table 3.2B Frequency of multidrug resistance (MDR) in Gram-positive pathogens causing BSI in Blantyre between 1998 and 2014

Year	*S. Pneumoniae*	*S .aureus*	*Enterococcus spp*	*Streptococcus spp.*
1998	85/335 (25.4)	45/99 (45.5)	18/26 (69.2)	18/90 (20.0)
1999	98/313 (31.3)	58/112 (51.8)	8/13 (61.5)	15/93 (16.1)
2000	80/238 (33.6)	75/156 (48.1)	3/8 (37.5)	21/143 (14.7)
2001	41/154 (26.6)	67/125 (53.6)	10/15 (66.7)	13/70 (18.6)
2002	75/231 (32.5)	59/102 (57.8)	2/7 (28.6)	8/72 (11.1)
2003	116/316 (36.7)	60/109 (55.0)	2/3 (66.7)	13/74 (17.6)
2004	83/265 (31.3)	47/98 (48.0)	5/7 (71.4)	22/79 (27.8)
2005	158/494 (32.0)	61/134 (45.5)	32/41 (78.0)	28/114 (24.6)
2006	110/413 (26.6)	25/76 (32.9)	6/13 (46.2)	15/70 (21.4)
2007	105/320 (32.8)	23/49 (46.9)	9/14 (64.3)	10/46 (21.7)
2008	54/160 (33.8)	25/72 (34.7)	12/14 (85.7)	6/29 (20.7)
2009	47/151 (31.1)	21/61 (34.4)	12/17 (70.6)	4/16 (25.0)
2010	34/119 (28.6)	18/72 (25.0)	10/13 (76.9)	13/28 (46.4)
2011	71/177 (40.1)	8/73 (11.0)	6/7 (85.7)	28/71 (39.4)
2012	30/106 (28.3)	18/104 (17.3)	15/16 (93.8)	7/41 (17.1)
2013	12/41 (29.3)	13/95 (13.7)	15/17 (88.2)	25/58 (43.1)
2014	12/31 (38.7)	11/94 (11.7)	15/18 (83.3)	16/53 (30.2)
2015	25/50 (50.0)	24/117 (20.50	29/34 (85.3)	19/46 (40.3)
2016	9/42 (21.4)	29/125 (23.20	37/48 (77.1)	8/31 (25.8)
Overall	**1211/3864 (31.3)**	**676/1873 (36.1)**	**246/331 (74.3)**	**289/1224 (23.6)**
p-value	**0.148**	**<0.001****	**<0.001****	**<0.001****

*=Decreasing trend; **=increasing trend

The surveillance also revealed 381 Gram-negative isolates that expressed co-resistance to ampicillin, cotrimoxazole, chloramphenicol, gentamicin, ciprofloxacin

and ceftriaxone. The most common pathogens with this phenotype included 122/381 (32.0%) *Klebsiella spp.,* 69 (18.1%) *E. coli*, 125 (32.8%) other Enterobacteriaceae and 68 (17.8%) *Acinetobacter spp.* isolates.

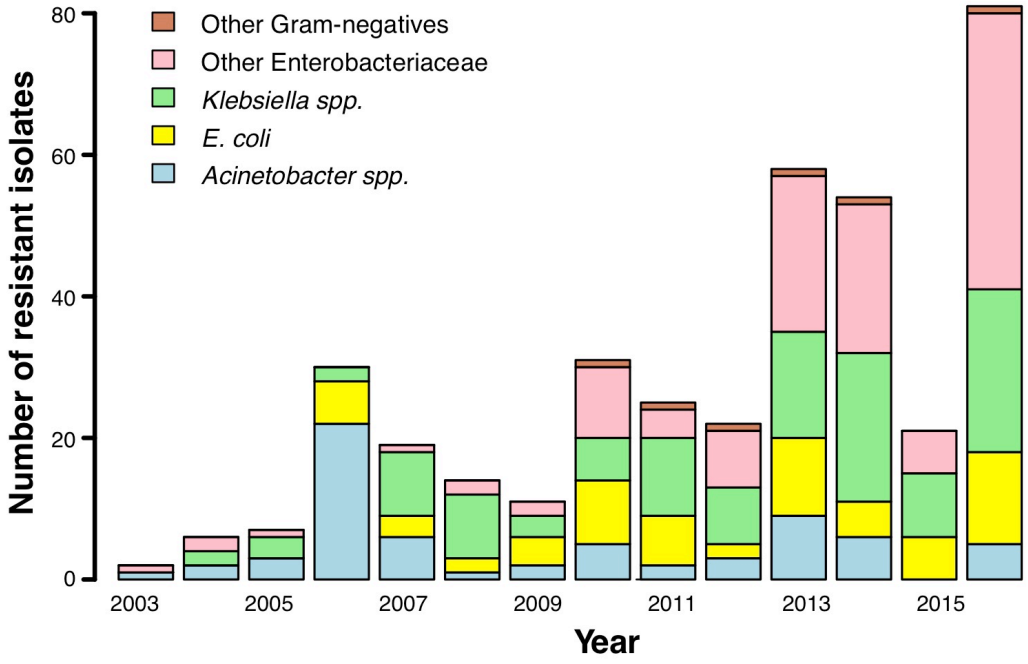

Figure 3.12 Distribution of isolates (by year and organism) that had expressed co-resistance to ampicillin, chloramphenicol, cotrimoxazole, ceftriaxone, ciprofloxacin and gentamicin between 2003 and 2014 in Blantyre.

The isolates with resistance to all the six locally available antimicrobial agents were first detected in 2003 with only 0.1% (2/2,372) BSI isolates identified, but this number has steadily increased such that by 2016, 7.3% (81/1,106) BSI isolates had expressed resistance to all the six locally available antimicrobial agents (Figure 3.12).

3.5. Discussion

The long term sentinel surveillance in Blantyre revealed that except for an on-going epidemic of *S*. Typhi, which emerged after 2010, there was a marked decline in the prevalence and incidence of BSI caused by all other pathogens between 1998 and 2016. This decline of BSI has occurred concurrently with the previously described interventions including extensive roll out of free antiretroviral therapy, PMCT, malaria control, improvements in food security, community management of malnutrition and the introduction of *Haemophilus influenzae* type b and PCV13 vaccines. Considerable reductions in mortality in children less than five years of age and adult HIV-related mortality in Malawi were also reported during the surveillance period (Feasey et al., 2014b; Zimba et al., 2012; Mwagomba et al., 2010). Given the association of bacterial BSI and factors such as HIV/AIDS, malaria and malnutrition and prominence of bacterial infection as a cause of death in these populations, it is likely that these trends are at least in part causally linked.

Whilst there was a declining trend in BSI, it is of major concern that resistance rates, especially to second-line antimicrobial agents, substantially increased in Blantyre. The surveillance revealed that ESBL and fluoroquinolone resistance was present in Blantyre even before ceftriaxone and ciprofloxacin were widely available, but their introduction was followed by marked increases in rates of resistance to these agents. The rates of ESBL and fluoroquinolone resistance of up to 18.7% and 20.6% in *E. coli* and 80.4% and 43.1% in *Klebsiella* reported here are very high, whether placed in an African or a global context (Dramowski et al., 2015; Castanheira et al.,

2014; Nakamura et al., 2012; Brink et al., 2012a). Moreover, in the other settings, ESBL and fluoroquinolone resistance have been reported to be more common in hospital acquired than community acquired infections, whereas the isolates in this study were from community acquired infection (Lautenbach et al., 2001; Rupp and Fey, 2003). Depending on the molecular mechanism of ESBL resistance, it is highly possible that the rates presented here are an underestimate of the true prevalence of ESBL-producing pathogens. Hence the need for further investigation beyond this PhD, which should include HA isolates.

Minimum incidence estimates suggest that *E. coli* and *Klebsiella spp.* BSI are particularly prominent in elderly patients followed by under five children, also consistent with results from other settings where children and the elderly were associated with the highest risk of *E. coli* BSI (Laupland, 2008). With respect to Blantyre, the Malawi National Statistical Office (NSO) estimates that life expectancy will increase from ~55 years in 2007 to ~70 years by 2030 (National statistical Office (NSO))Increased life expectancy may increase the pool of persons at risk of *E. coli* and *Klebsiella* BSI, with the attendant risk of drug resistance increasing the prominence of these pathogens with time.

Resistance to first line antimicrobials fluctuated. In some Enterobacteriaceae, RFL rates had begun to decline, primarily due to less chloramphenicol resistance. The decline in isolates with chloramphenicol resistance was most pronounced in *E. coli,* but the molecular explanation of the decline remains uncertain. Overall, however, this partial re-emergence of chloramphenicol susceptibility is not sufficiently great

to permit its reintroduction as an agent for the empirical management of sepsis in Blantyre, especially in the context of an epidemic of MDR Typhoid fever.

Chloramphenicol and penicillin have been commonly used in combination for the empirical management of sepsis in Malawi for many years (Everett et al., 2011). It was interesting to note that almost all (99.0%) of the *S. pneumoniae* isolates were still susceptible to this combination despite its wide usage, although this might be due to the complex multistep process required to confer β-lactam resistance (Hakenbeck et al., 2012). It is equally important to note that penicillin resistance has started to increase since the introduction of PCV13 in 2011. PCV13 introduction has been associated with a general decline in penicillin resistant *S. pneumoniae* in South Africa (von Gottberg et al., 2014). However, increasing prevalence of penicillin resistant *S. pneumoniae* serotype 19F and non-vaccine serotypes such as 19A and 15A have been reported following the introduction of the PCV7 and PCV13 outside SSA (Richter et al., 2009; Nakano et al., 2016). The increase in the proportion *S. pneumoniae* that are penicillin resistant at a time when *S. pneumoniae* BSI is in decline raises the possibility that there has been a change in serotype distribution following vaccine introduction as has been the case in the other settings (Nakano et al., 2016; Richter et al., 2009).

MRSA currently remains an infrequent cause of BSI in Blantyre. Prevalence of MRSA amongst the *S. aureus* isolates was similar to proportions from countries such as Mozambique and Zimbabwe, but much lower to those reported in South Africa (Mandomando et al., 2010; Mauchaza et al., 2016; Perovic et al., 2015). The

difference in prevalence of MRSA amongst S. aureus may be a reflection of the fact that cultures described in this thesis were taken from community admissions to medical wards; while nosocomial infections are yet to be studies in in depth in Blantyre. It might also reflect that unlike in South Africa, medical devices placing patients at risk of MRSA BSI such as central venous catheters are uncommonly used in low-income countries such as Malawi. However, sustained presence of MRSA as a low-level cause of CA-BSI in Blantyre is still of considerable concern as its relative importance as a BSI pathogen may greatly change if surveillance expands to cover surgical patients or nosocomial infections or if medical practice changes, for example a renal dialysis unit is under development.

3.6 Limitations

It is possible that CA BSI was missed if persons died at home or were not referred to QECH, hence the rates reported in this chapter are considered to be minimum estimates. It is also possible that some cases were misclassified as CA when they were nosocomial infections, due to the lack of some patient's hospitalisation history. Because it is uncommon for patients to have a follow-up blood culture, it was unlikely that the surveillance captured much nosocomial infection.

ESBL screening was not introduced until 2003, and as such ESBL-producing pathogens may have been circulating, but undetected before then. ESBL screening by cefpodoxime disc testing was not introduced until 2007, consequently some

isolates may have been falsely classified as ESBL producing prior to 2007. In the case of *S. aureus*, cefoxitin screening for MRSA replaced methicillin in 2010, although the small increase in sensitivity gained will have made minimal difference to the reported findings.

3.7 Conclusion

Overall declines in bacterial BSI have been accompanied by a rise in antimicrobial resistance in all bacterial BSI pathogens at QECH, especially in Gram-negative organisms and the emergence of methicillin resistance in *S. aureus* and increased pneumococcal resistance following vaccine introduction. Ceftriaxone and ciprofloxacin have been essential for the management of bacterial BSI in a context where human immunosuppression and bacterial multidrug resistance are highly prevalent. The emergence of ESBL, fluoroquinolone and gentamicin resistance and MRSA highlight the growing challenge of BSI that is effectively impossible to treat in this resource-limited setting

Chapter four: Genomic landscape of extended antimicrobial resistance in *Escherichia coli* from Blantyre, Malawi

4.1 Overview

Efforts to treat *E. coli* infections are increasingly being compromised by the rapid global spread of AMR. ESBL producing *E. coli,* which globally are dominated by the sequence ST131, are a major health concern. Whilst AMR in *E. coli* has been extensively investigated in resource rich settings, in SSA, molecular patterns of AMR in *E. coli* are not well described. The aims of the work in this chapter were:

- Describe the population structure of *E. coli* isolates in Blantyre, Malawi.
- Identify the molecular determinants of AMR in *E. coli* isolates from Blantyre
- Determine the *E. coli* lineages associated with ESBL and fluoroquinolone resistance

Analysis of genome sequences of *E. coli* isolates from Blantyre demonstrated that the *E. coli* population in Malawi is very diverse and isolates have genomes that are highly plastic. CTX-M-15 was identified as the most common ESBL type whereas fluoroquinolone resistance was largely due to chromosomal mutations in the *gyrA* gene. Unlike in other settings where a limited number of *E. coli* STs, mostly ST131,

are associated with the current globally dominant ESBL type, CTX-M-15, in Malawi, a diverse number of unrelated STs are CTX-M-15 producing and no dominant CTX-M-15 ST was identified.

4.2 Introduction

E. coli was the second most common Gram-negative cause of CA-BSI in Blantyre accounting for 8.8% between 1998 and 2016 (see Chapter Three, Section 3.4.2.1). The majority of *E. coli* isolates were MDR to first-line agents including ampicillin, cotrimoxazole and chloramphenicol. However, more worrying is the fact that there had been a significantly increasing incidence of *E. coli* isolates that are resistant to the last resort antimicrobial agents of ceftriaxone, ciprofloxacin and gentamicin. Third generation cephalosporins and fluoroquinolones are the last resort agents in Malawi hence *E. coli* expressing resistance to these agents cause infections that are effectively untreatable in this setting (See Chapter Three, section 3.5). Within SSA and indeed in many parts of the world, ESBL producing and fluoroquinolone resistant *E. coli* have become a major threat to public health (WHO, 2014).

Global studies have identified CTX-M derivatives, and in particular CTX-M-15, to be the dominant and most widely disseminated ESBL type produced by *E. coli* isolates (Nicolas-Chanoine et al., 2008). WGS studies have further shown that despite the high genetic diversity associated with the *E. coli* species (Chaudhuri and Henderson, 2012), ESBLs especially of the CTX-M family are strongly associated with specific clonal *E. coli* lineages (D'Andrea et al., 2013). In the case of CTX-M-15, it is predominantly associated with a globally disseminated ST131 clone (Matsumura et

al., 2015; D'Andrea et al., 2013). The ESBL producing lineages are also associated with fluoroquinolone resistance (Peirano et al., 2010), which for *E. coli*, is mainly conferred by chromosomal mutations in the quinolone binding target genes such as *gyrA* gene (Hooper and Jacoby, 2015b). Recently, plasmid mediated genes such as *qnr*, *qep* and *aac(6")-Ib-cr* have also been shown to confer low-level fluoroquinolone resistance (Hooper and Jacoby, 2015b). Both cephalosporins and fluoroquinolones are the antimicrobial agents of choice for treating severe bacterial infection in Malawi and many other countries in SSA (Feasey et al., 2015c; Leopold et al., 2014). Whilst the emergence of ESBL producing and fluoroquinolone resistant *E. coli* is becoming major public health concern in SSA, there have been limited studies to provide insight into the molecular patterns underlying the emergence, spread of ESBL and fluoroquinolone resistance in the region. This chapter presents the first *E. coli* WGS study in Malawi, which provides a baseline for understanding the population structure of *E. coli* and identifies *E. coli* genotypes and lineages associated with ESBL and fluoroquinolone resistance.

4.3 Sequenced Malawian *E. coli* isolates

Ninety-three isolates identified as *E. coli* at MLW using the culture methods outlined in Chapter Two (section 2.2.2) were selected for WGS on the basis of their phenotypic AMR profiles, year and source of isolation. Fifty-seven (57/93[61.3%]) were from blood culture, of which 20 (35.1%) were from adults, 29 (54.1%) from children and 8 (14.0%) isolates were from patients whose age was not known. Twenty (21.5%) isolates were from CSF, of which three (15.0%) were from adults,

nine (45.0%) from children and eight (40.0%) from patients with unknown age (Table 4.1). The collection also included 16 (16.2%) rectal swab isolates, all from adult patients.

Table 4.1 Distribution of *E. coli* isolates selected for WGS by clinical source of isolation and age group

	Clinical source (% of total isolates)			
Age group	BC	CSF	RS	**Total**
Adult	20 (21.5)	3 (3.2)	16 (17.2)	**39 (41.9)**
Children	29 (31.2)	9 (9.7)	0 (0.0)	**38 (40.9)**
Unknown	8 (8.6)	8 (8.6)	0 (0.0)	**16 (17.2)**
Total	**57 (61.3)**	**20 (21.5)**	**16 (17.2)**	**93 (100)**

Blood and CSF samples were collected over a period spanning 1996-2014, whereas all samples for the rectal swab isolates were all collected in 2009 (Figure 4.1). Susceptibility profiles to the six antimicrobial agents tested for each selected isolate are shown in the Appendix 4.

Following the sequencing of the 93 *E. coli* genomes, Kraken (Wood and Salzberg, 2014) was used to confirm the classification of the isolates to species level based on WGS matching. The characterisation by Kraken identified three isolates in this *E. coli* collection to be *K. pneumoniae* and these were transferred to the *K. pneumoniae* dataset (Described in Chapter Five). Two isolates were respectively identified as *C. rodentium* and *P. mirabilis* and were excluded from further analysis. Two genome sequences with very high (7,257,854 bp and 10,508,894 bp) and one with very low

(426 bp) genome sizes (Appendix 3) that were uncharacteristic of *E. coli* (genome size ~5MB) were also excluded from further analysis. Nine *E. coli* genome sequences were identified from the *K. pneumoniae* genome collection (Chapter Five, section 5.4) and were analysed together with the remaining 85 isolates to have a total of 94 genome sequences for analysis. The disconcordance between the culture and *in silico* classification could possibly be as a result of mislabelling or contamination of samples.

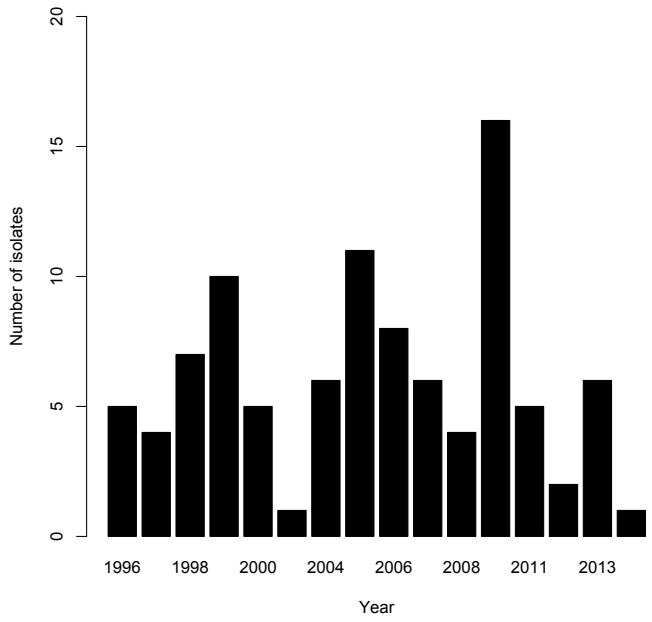

Figure 4.1 Distribution of *E. coli* isolates selected for WGS by year of isolation

4.4 Genetic diversity and population structure of Malawian *E. coli* isolates

4.4.1 Phylogroups and multi-locus sequence type diversity

Characterisation by *in silico* PCR showed that the *E. coli* isolates belonged to five different phylogroups namely A, B1, B2, D and F. Phylogroup B2, was the largest phylogroup in this collection (Figure 4.2B) containing 34.0% (32/94) of the isolates. Phylogroup A was the second largest phylogroup with 23.4% (22/94) isolates. Other Phylogroups included D, 20.4% (19/93) of isolates, F, 14.0% (13) and B1, 7.5% (7/93) of isolates (Figure 4.2B).

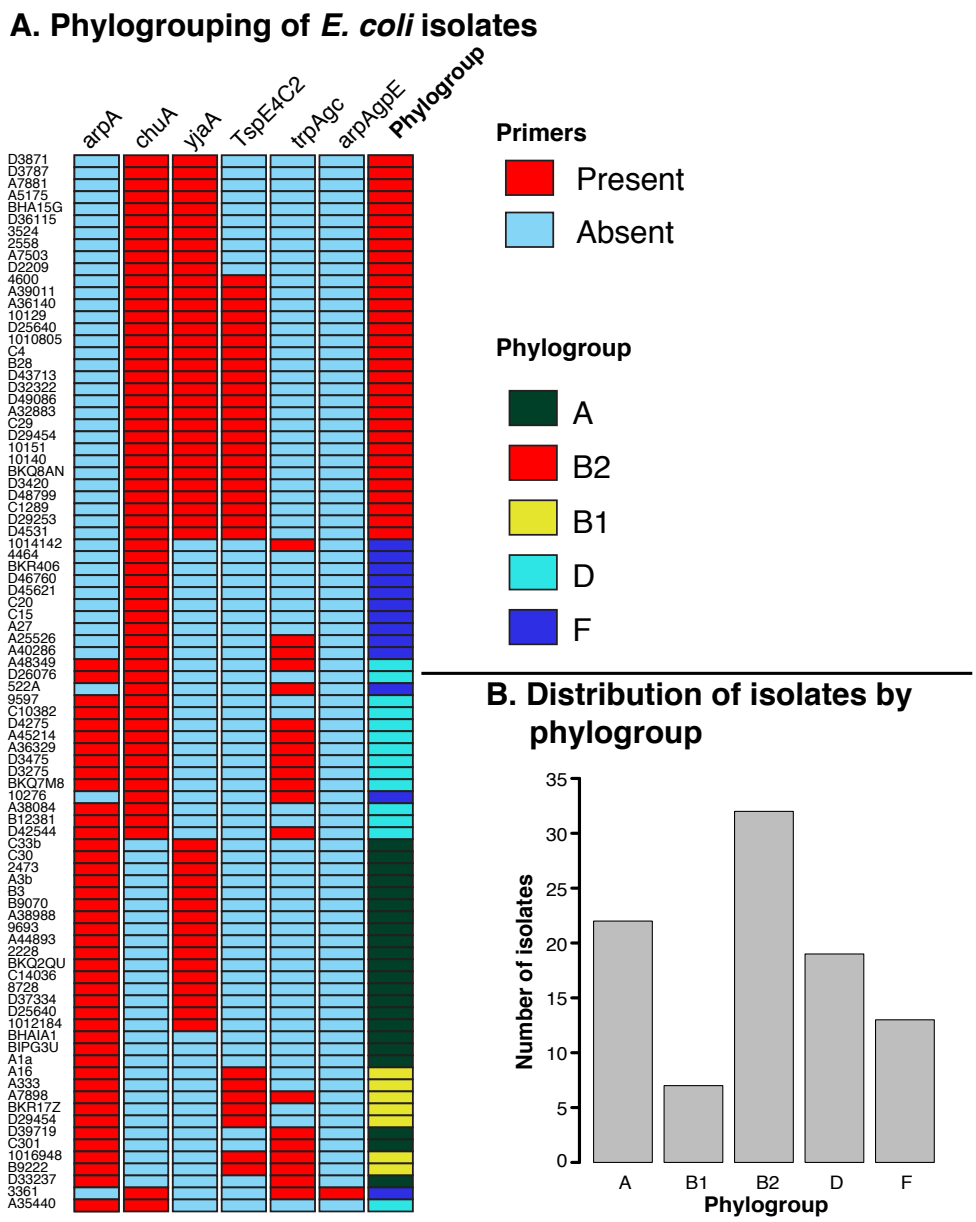

Figure 4.2 Determination of *E. coli* phylogroups. Determination of *E. coli* Phylogroups based on absence or presence of sequences of primers *arpA, chuA, yjaA, TsPE4C2, trpAgc, arpAgpE* in whole genome assembly sequences of Malawian *E. coli* isolates and distribution of isolates by phylogroup.

In silico MLST characterised 89 *E. coli* isolates into 43 different STs while five isolates could not be uniquely typed as they bore previously unreported alleles. ST131 was the most common sequence ST identified with 15.1% (14/93) of the isolates and was followed by ST12, which was identified with 10.8% (10/94) of the isolates. Other common STs included ST69 (5.0%), ST10 (5.0%) and ST391 (4.0%) (Table 4.2). The majority (65.1%) of the STs however, were singletons. Excluding isolates with unknown STs, the Simpson diversity index was calculated to be 0.94. This diversity index is high and does suggest a high likelihood that any two *E. coli* isolates randomly sampled from this population would belong to two different STs.

Table 4.2 Distribution of Malawian *E. coli* isolates by ST

ST	Phylogroup	No. of isolates	%	$]({}^{n}/_{N})^{2}$; N=89
ST131	B2	14	15.1%	0.02
ST12	B2	10	10.8%	0.01
ST73	B2	3	3.2%	0.001
ST636	B2	2	2.2%	0.0005
ST10	A	5	5.4%	0.003
ST167	A	3	3.2%	0.001
ST617	A	2	2.2%	0.0005
ST48	A	1	1.1%	0.0001
ST52	A	1	1.1%	0.0001
ST120	A	1	1.1%	0.0001
ST361	A	1	1.1%	0.0001
ST399	A	1	1.1%	0.0001
ST410	A	1	1.1%	0.0001
ST652	A	1	1.1%	0.0001
ST1286	A	1	1.1%	0.0001
ST1312	A	1	1.1%	0.0001
ST4358	A	1	1.1%	0.0001
ST4656	A	1	1.1%	0.0001
ST69	D	5	5.4%	0.003
ST391	D	4	4.3%	0.002
ST38	D	2	2.2%	0.0005
ST1567	D	2	2.2%	0.0005
ST5148	D	2	2.2%	0.0005
ST280	D	1	1.1%	0.0001
ST448	B1	2	2.2%	0.0005
ST196	B1	1	1.1%	0.0001
ST223	B1	1	1.1%	0.0001
ST500	B1	1	1.1%	0.0001
ST977	B1	1	1.1%	0.0001
ST1084	B1	1	1.1%	0.0001
ST15	B2	1	1.1%	0.0001
ST504	B2	1	1.1%	0.0001
2279	B2	1	1.1%	0.0001
648	F	3	3.2%	0.001
405	F	2	2.2%	0.0005
59	F	1	1.1%	0.0001
62	F	1	1.1%	0.0001
335	F	1	1.1%	0.0001
354	F	1	1.1%	0.0001
362	F	1	1.1%	0.0001
501	F	1	1.1%	0.0001
1163	F	1	1.1%	0.0001
2141	F	1	1.1%	0.0001
Simpson's index, $D = \sum({}^{n}/_{N})^{2}$				**0.0533**

$$Simpson's\ diversity\ index = I - D$$

$$= 1 - 0.05$$
$$= 0.95$$

When the phylogroup and MLST data were mapped to the *E. coli* phylogeny, there was concordance between phylogroups and the SCs. (Figure 4.7), with the exception of phylogroup F, which clustered with isolates from other phylogroups in multiple SCs. The phylogenetic clustering of phylogroup F isolates with isolates in other phylogroups is not unique to this dataset. Evidence of lack of concordance between clustering of *E. coli* isolates phylogroup F and phylogeny appears in a study by Clermont *et al.* where a number of phylogroup F isolates cluster with isolates in phylogroups B2 and D instead of the main phylogroup F cluster (Clermont et al., 2013). Invasive *E. coli* isolates are mostly from phylogroups B2 and D (Croxen et al., 2013). However, it was noted that isolates in this collection were clustered independent of their status of invasiveness.

Except for EC_SC2, which consisted of ST131, and EC_SC5, which consisted of ST12 and a few other STs, the rest of the other SCs comprised of multiple diverse and distantly related ST. EC_SC1 comprised of nine STs, EC_SC3 comprised of 23 STs in SC3 while EC_SC4 comprised of six STs.

4.4.2 Pan and core genomes

Automated annotation of the 94 confirmed *E. coli* predicted a mean of 4,845 coding sequences (CDS) per genome for a mean genome length of 5.2MB. Roary pan genome analysis pipeline identified a pan genome consisting of 22,934 non-orthologous genes. The core genome (set of genes in ≥99 genomes) comprised of 2,210 genes, encoded on ~1.7MB. This core genome therefore, represents 9.6% (2,210/22,934) of the pan genome, 45.6% of coding region per genome and 32.7% of the mean genome length. The accessory genome (set of genes identified in < 99 genomes) of isolates comprised of 20,724 non-orthologous genes i.e. 90.4% of the pan genome size. A substantial proportion (8,362/20,724 [40.3%]) of the accessory genes were singletons. The number of genes in the pan genome, core genome, set of unique genes and set of new gene per additional genome was iteratively calculated for 10 different sets of genome sequences at different sample sizes ranging between 1 and 94. A graphical analysis showed that the pan genome and number of isolate specific genes were continuously increasing per each additional genome sequence. On the other hand the core genome and number of new genes declined with additional genome sequences (Figure 4.3).

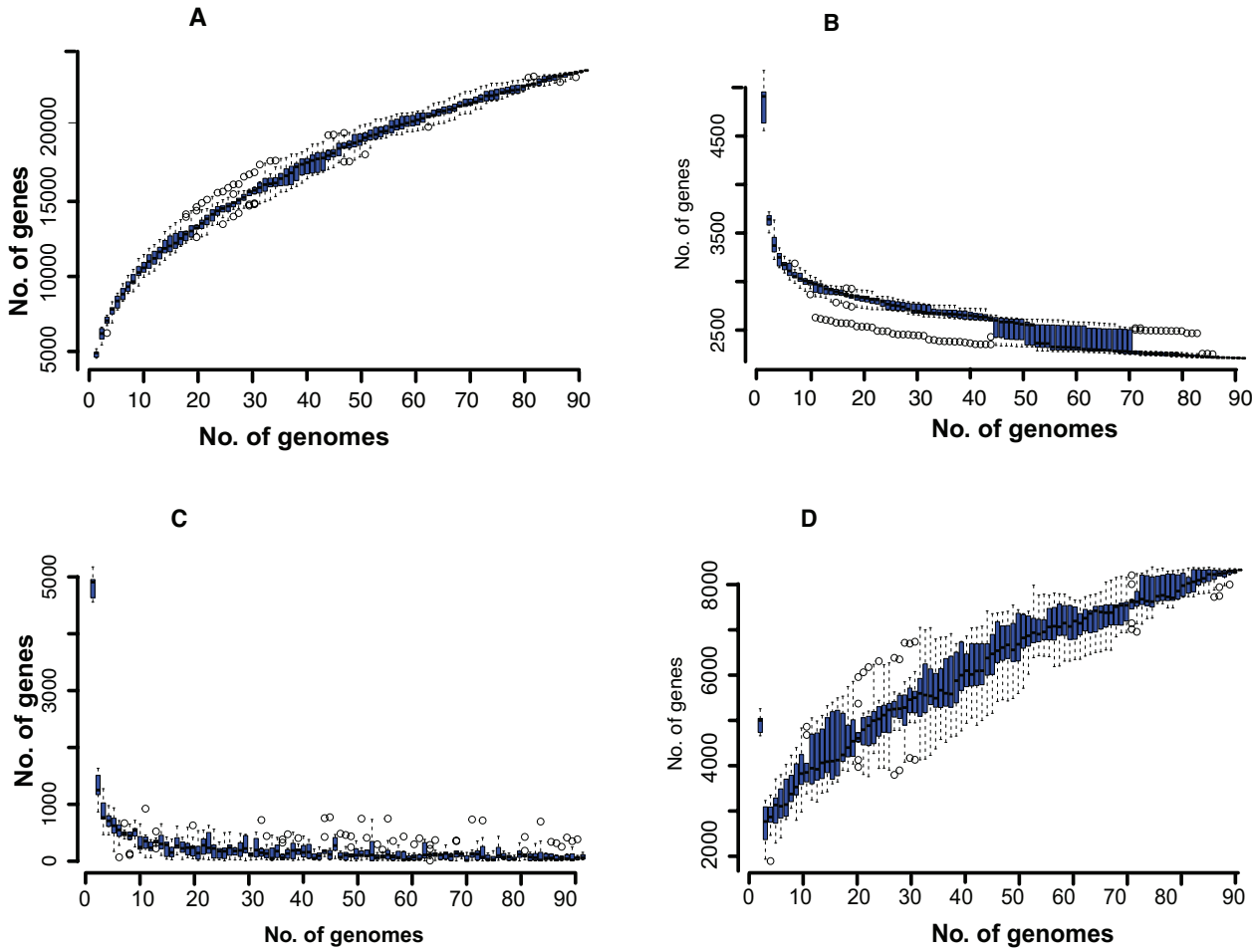

Figure 4.3 Distributions of genes in the *E. coli* isolates collected from Blantyre.

Box plots showing distribution of number of genes in 10 samples obtained iteratively for a given number of genome sequences. (A) Total number of non-orthologous genes in the pan genome of given (B) Number of core genes per given number of genomes sequences(C) Number of new genes identified per added genome sequence (D) Number of unique genes per given the number of genomes.

E. coli has been reported in a number of studies to have an open pan genome i.e. genomic analyses of new genome sequences results in identification of new genes. To determine whether the increasing pan genome curve was bounded, the expansion of pan genome of the Malawian *E. coli* isolates was modelled using the power-law regression equation, which take the general form $N = \beta g^{\gamma}$, where $N = pan\ genome\ size$ and $g = number\ of\ genome\ sequences$, as described in chapter Two).

The analysis revealed a strong log-linear relationship ($R^2 = 0.99; p < 0.0001$) between the pan genome size and number of genome sequences included for analysis. The parameters β and γ in the power-law regression equation $N = \beta g^{\gamma}$ (were estimated by the log-linear regression equation:

$$\ln(N) = \ln(\beta) + \gamma \ln(g)$$

Results of the regression analysis of the above equation yielded the following values when fitted with the data on number of genes in the pan genome:

$\ln(\beta) = 8.44 \Rightarrow \beta = e^{8.4}$;

$\gamma = 0.4$;

and power-law regression equation describing the expansion of the *E. coli* pan genome cumulative curve (Figure 4.4) was defined by the equation:

$$N = e^{8.4} g^{0.4}$$

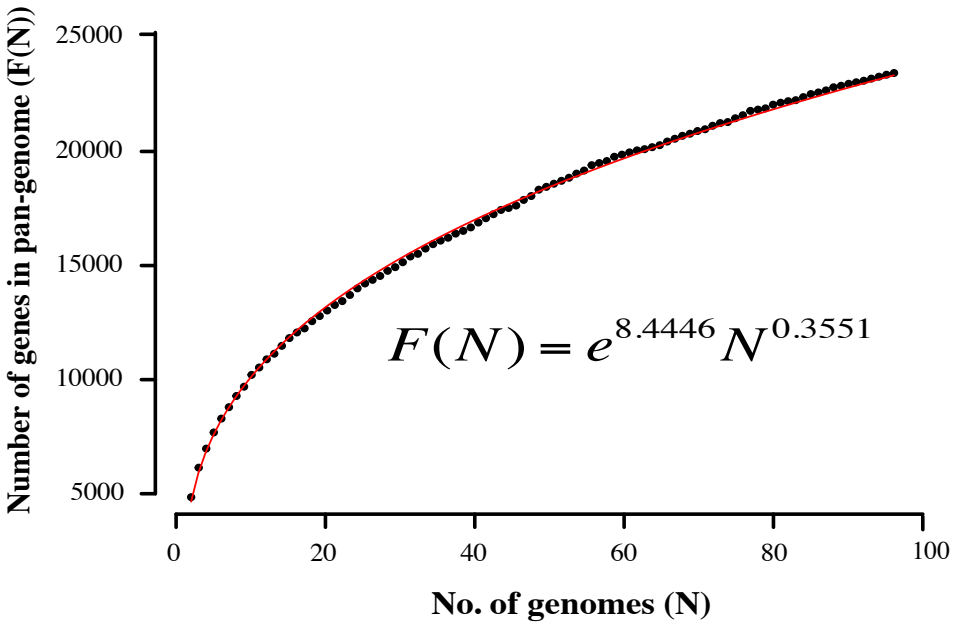

Figure 4.4 Expansion of the Malawian *E. coli* pan genome with increasing number of genome sequences as modelled by the power law-regression equation $F(N) = \beta N^{\gamma}$**.** The curve in red illustrates the expansion of the E. coli pan genome as modelled by the regression equation. The points in black represent the observed average pan genome size from a given number of genomes (N).

Since $\gamma\ (= 0.4) > 0$, the curve defined by the equation $F(N) = e^{8.4} g^{0.4}$ is unbounded, indicating that the *E. coli* population in Blantyre has an open pan genome consistent with previous studies elsewhere (Touchon et al., 2009; Snipen et al., 2009). The results of this pan genome analysis do indicate that genomes of *E. coli* isolates in Malawi have a large variable region and the ability to easily acquire

foreign DNA. The ability to easily acquire genes has enabled *E. coli* to easily adapt to heterogeneous environments and acquire new phenotypic traits such as resistance to antimicrobial agents and virulence.

4.4.3 Population structure of Malawian *E. coli*

The structure of the *E. coli* population in Malawi was investigated using two approaches: the Bayesian population structure analysis and phylogenetic analysis using the hierBAPS module in BAPS and phylogenetic analyses. *E. coli* phylogroups were also determined and examined in relation to the population structure analyses.

4.4.3.1 Bayesian population structure of Malawian E. coli isolates

Short paired-end read sequences of the Malawian *E. coli* genomes were mapped to the genome sequences of the *E. coli* reference strain MG1665 (Accession number CP00096) and generated a multiple alignment. hierBAPS was run on the alignment after removal of predicted recombination regions by Gubbins . This BAPS analysis resulted in the identification of five sequence clusters (SC), designated here as EC_SC1, EC_SC2, EC_SC3, EC_SC4 and EC_SC5.

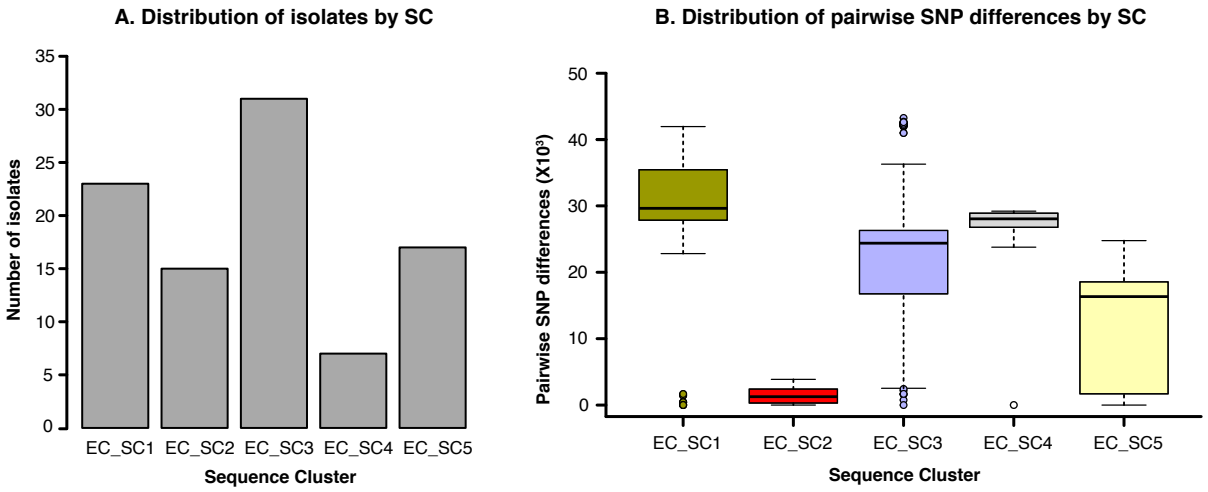

Figure 4.5 Distribution of isolates and pairwise SNP differences by SC. (A) Bar chart of number of isolates per SC and (B) Boxplot of pairwise SNP differences by SC

EC_SC3 had the highest number of isolates, followed by EC_SC1 and EC_SC5, while EC_SC4 was the cluster with the least number of isolates (Figure 4.5A). A core genome alignment was generated by concatenating the multiple alignments of the core genes. Calculation of core genome pairwise SNP differences showed heterogeneous nucleotide sequence diversity amongst the SCs. Overall, means of pairwise SNP differences per SC ranged between 1,419 SNPs to 29,010 SNPs. EC_SC2 had the lowest mean pairwise SNP difference of 1,419 SNPs, and followed by EC_SC5 as the second least diverse SC with a mean pairwise SNP difference of 12,300 (Figure 4.5B). Much higher sequence diversity was exhibited by the other SCs, with EC_SC1 having the highest mean pairwise SNP difference of 29,010 SNPs while EC_SC3 and EC_SC4 followed with mean pairwise SNP differences of 22,230 and 23,950 core genome SNPs, respectively (Figure 4.5B).

4.4.3.2 Phylogeny of Malawian *E. coli* isolates

The phylogeny of the *E. coli* isolates was first reconstructed using two sequence alignments: the core genome alignment and the alignment obtained from mapping short-read sequences of the *E. coli* genomes to the genome sequence of the reference strain MG1665 and adjusting for recombination. The objective here was to determine if the topologies of the phylogenetic trees constructed using the two approaches would be different. The two most divergent core genome sequences were separated by 191,504 SNPs. The phylogenetic tree constructed from the alignment of short read sequences to reference strain was constructed based on 139,769 SNP sites. A tanglegram comparison of the two phylogenetic trees drawn with Dendroscope revealed a similar topology between the two trees. However, some branches on the phylogenetic tree generated from mapping alignment had shorter branches than corresponding branches on the core genome tree (Figure 4.6). This difference in the branch lengths reflects the higher number of SNP sites in the core genome alignment than in the recombination free mapping alignment.

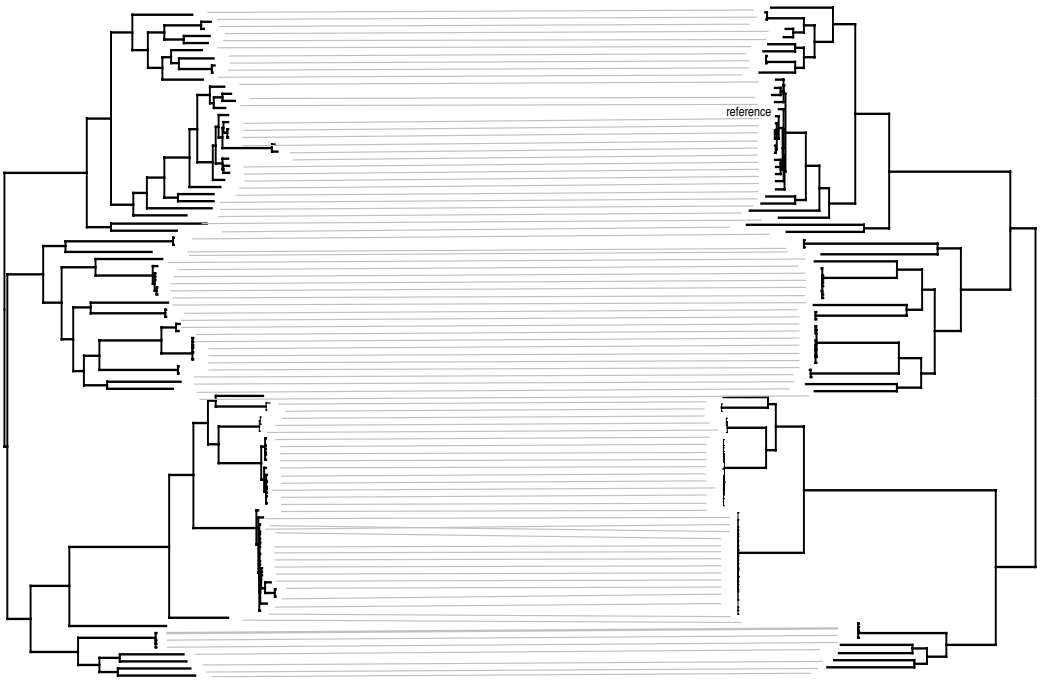

Figure 4.6 Tanglegram comparing ML phylogeny based on core genome alignment (right) with ML phylogeny based on whole genome sequence SNP alignment (from mapping).

The differences in lengths between corresponding branches in some clades of core genome phylogeny and the phylogeny from mapping alignment could be due to the higher SNP density in core genome, probably occurring because the isolates are distantly related rather than due to recombination, might have been predicted as recombination sites by Gubbins and removed from the alignment. In this case, the alignment on which Gubbins was applied would have fewer SNP sites. Considering that over 50.0% of the mean length of the *E. coli* genomes included contain variable regions (Section 4.5.1), using an alignment adjusted for recombination

underestimates the divergence of the genomes. Hence, the core genome tree was used to model the phylogeny of the entire collection, where as trees obtained from mapping alignments have been used for lineage specific analyses.

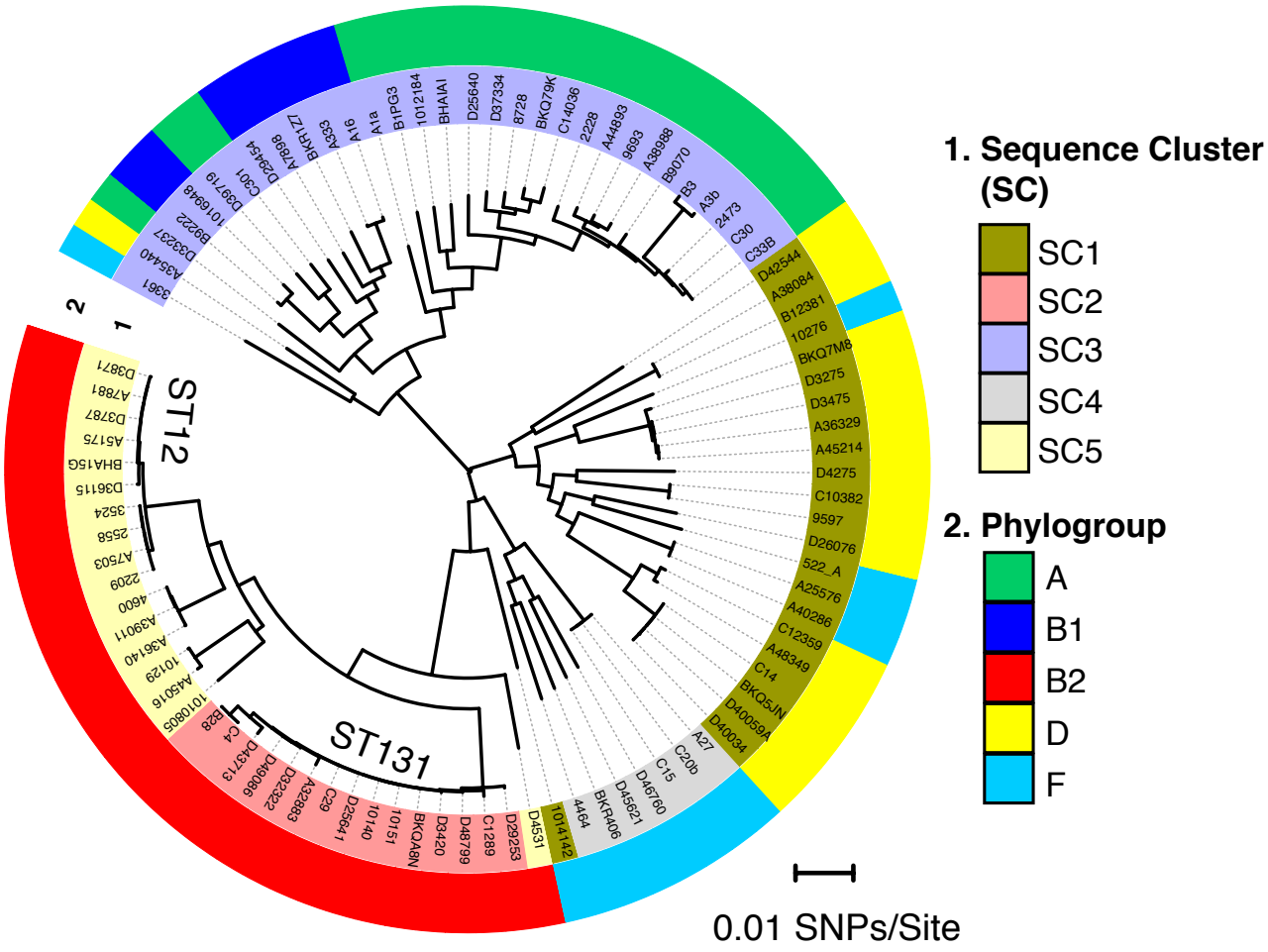

Figure 4.7 Population structure Malawian *E. coli* isolates collected from Blantyre, Malawi. Circular ML core genome phylogenetic tree of Malawian E. coli isolates rooted mid-point of the longest branch separating the two most divergent isolates. The inner ring designates identified SCs by hierBAPS and the outer ring designates phylogroups identified by *in silico PCR*.

The BAPS clustering data were mapped to the core genome phylogeny and this revealed concordance between phylogenetic clustering and the SCs determined by BAPS clustering (Figure 4.7). Consistent with the within cluster sequence diversity, EC_SC2 and EC_SC5 were the most clonal lineages, where as EC_SC1, EC_SC3 and EC_SC4, which were characterised by high sequence diversity, consisted of long branches in distantly related sub-clades.

4.4.4 Malawian ST131 isolates in context of global population structure of ST131

ST131 is perhaps the most important and widely studied *E. coli* ST, ref;ecting its global distribution and AMR and virulence profiles (Nicolas-Chanoine et al., 2014). To place the Malawian ST131 isolates in the context of other global ST131 isolates, 17 isolates from a collection that was used to describe a global population structure of *E. coli* were selected. The global population structure of ST131 had identified three sub-lineages, designated A, B and C constituting the *E. coli* ST131 lineage (Petty et al., 2014). The 17 isolates included in this analysis were selected from all the three clades. Paired-end read genome sequences of the selected isolates were retrieved from ENA and together with the 14 ST131 isolates identified in the Malawian *E. coli* collection mapped to ST131 EC958 (Accession number HG841718) reference genome sequence.

A ML phylogeny, reconstructed from the alignment after recombination sites were predicted and removed by Gubbins, and BAPS analysis revealed four clusters. The first three clusters were the ST131 sub-lineages A, B and C, previously identified

from the global collection. All the selected global ST131 isolates clustered into these clades as expected (Figure 4.8). However, only three of the Malawian ST131 isolates clustered into these previously known ST131 sub-lineages. Two isolates (B28 and C4) clustered with isolates in the ST131 sub-clade C and one isolate (D29253) clustered with isolates in ST131 sub-clade B. The rest of the isolates formed a cluster, designated here as D, that was unique to the Malawian ST131 isolates, branching off from sub clade C, suggesting common ancestry with this globally disseminated CTX-M-15 clade.

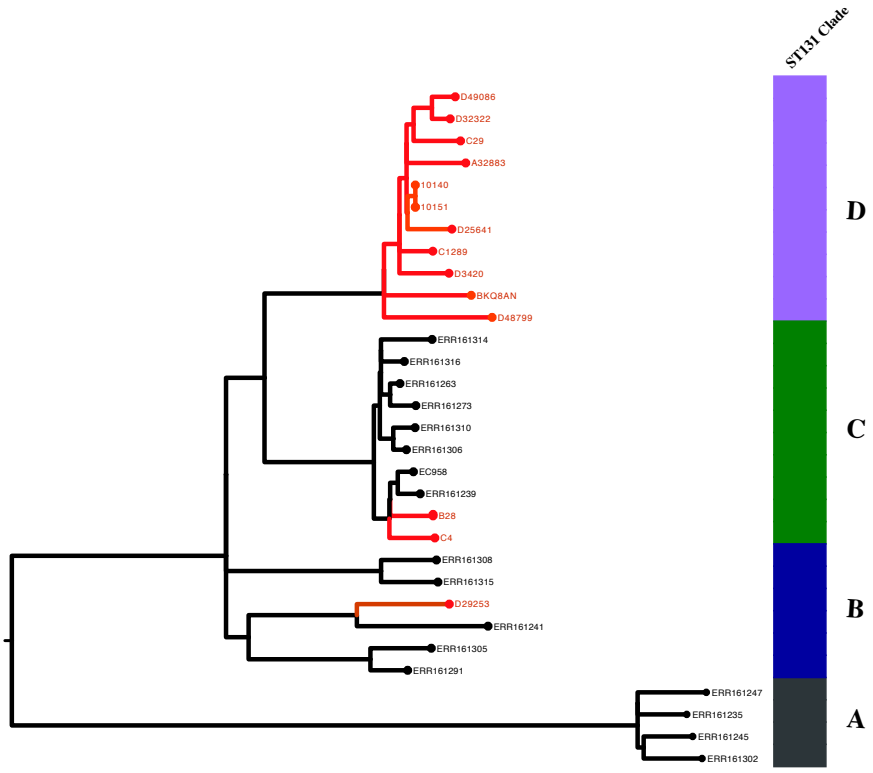

Figure 4.3 Malawian ST131 isolates in a global context. A phylogeny illustrating evolutionary relatedness of Malawian *E. coli* ST131 isolates (red) and global *E. coli* ST131 isolates (black). *E. coli* ST131 clades A-C are as defined (Petty et al., 2014), while clade D is unique Malawi collection and identified from this work.

4.5 Genetic determinants of antimicrobial resistance in *E. coli* genomes

A total of 25 acquired AMR genes were identified through a BLAST search against the ResFinder database. The identified genes included those that confer resistance to commonly available antimicrobial agents such as aminopenicillins, cephalosporins, aminoglycosides, fluoroquinolones, sulphonamides, trimethoprim, macrolides and tetracyclines.

Table 4.3 Distribution of AMR genes among Malawian *E. coli* isolates

Gene	**Known AMR phenotype**	**No. of Isolates(%)**
sul2	Sulphonamides	88 (93.6)
strA	Aminoglycosides	87 (86.2)
strB	Aminoglycosides	87 (92.6)
dfrA	Trimethoprim	86 (91.5)
$bla_{\text{TEM-1}}$	β-Lactams	74 (78.7)
catA	Chloramphenicol	61 (64.9)
sul1	Sulphonamides	51 (54.3)
tetB	Tetracycline	51 (54.3)
aac3-IIa	Aminoglycosides	37 (39.4)
tetA	Tetracycline	38 (40.4)
mphA	Macrolides	35 (37.2)
$bla_{\text{CTX-M-15}}$	ESBL	20 (21.3)
$bla_{\text{OXA-1}}$	β-Lactams	20 (21.3)
aac(6')-lb-cr	Aminoglycosides, fluoroquinolone	18 (19.1)
aph3	Aminoglycosides	15 (16.0)
qnrB1	Fluoroquinolones	4 (4.3)
floR_2	Chloramphenicol	3 (3.2)
$bla_{\text{SCO-1}}$	β-Lactams	3 (3.2)
aadA2	Aminoglycosides	2 (2.1)
$bla_{\text{SHV-12}}$	ESBL	2 (2.1)
catB	Chloramphenicol	2 (2.1)
cmlA1	Chloramphenicol	1 (1.1)
$bla_{\text{OXA-10}}$	ESBL	1 (1.1)
$bla_{\text{SHV-1}}$	β-Lactam	1 (1.1)
qepA	Quinolones	1 (1.1)

4.5.1 Presence of AMR genes and phylogenetic clustering

When the AMR phenotype and AMR genes were overlaid on the *E. coli* phylogeny, there was a strong association between some of the identified AMR genes and the available AMR phenotypes (Figure 4.9). Both AMR phenotypes and genes were distributed independent of phylogenetic clustering of the host isolates. Similarly, there were no specific AMR phenotypes or genotypes associated with isolates from a particular source or patient age group (Figure 4.9).

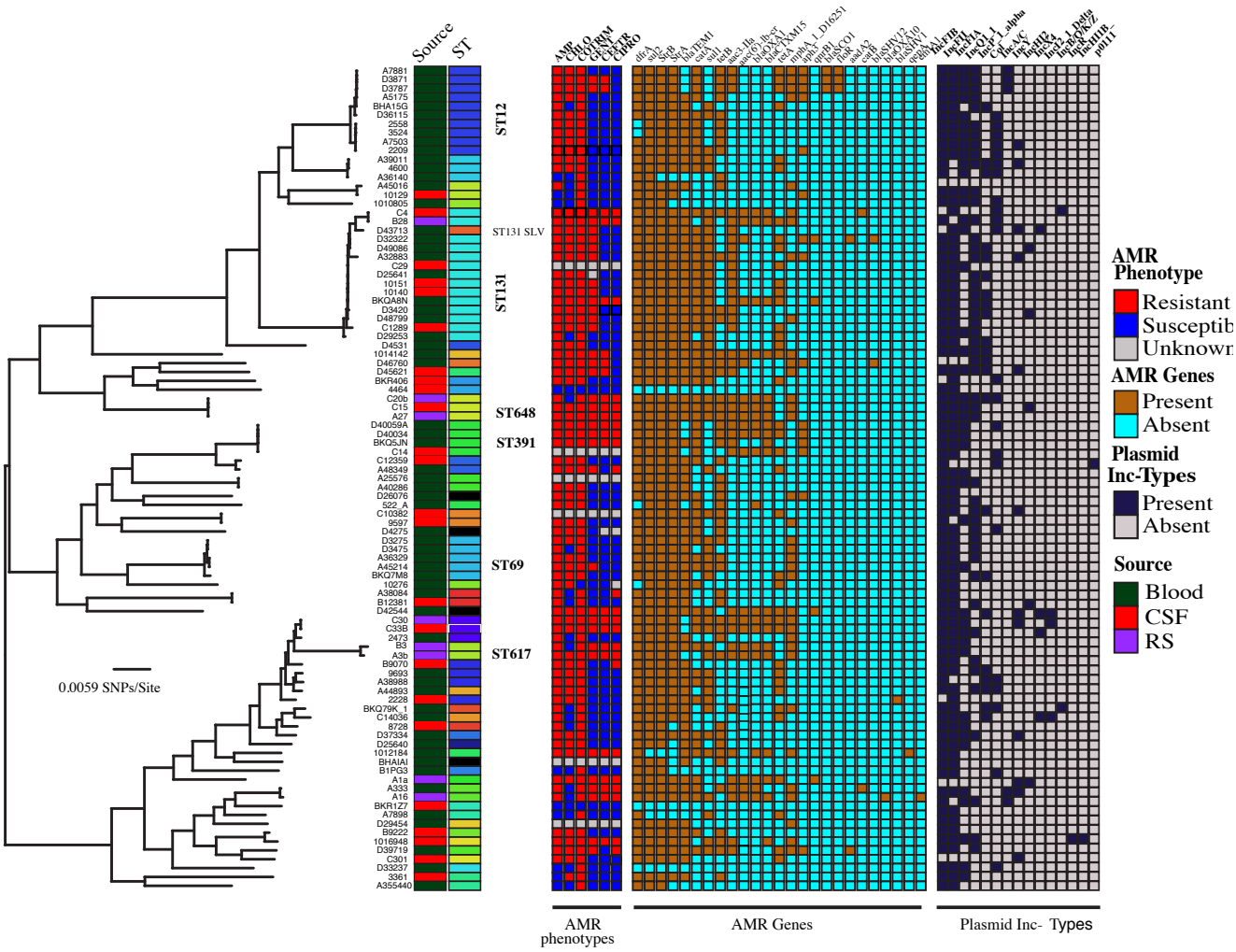

Figure 4.4 Distribution of AMR phenotype profiles, acquired AMR genes and plasmid incompatibility groups across the phylogeny of Malawian *E. coli* isolates.

4.5.2 Genetic determinants of β-Lactam resistance

Seven genes associated with resistance to β-lactam antimicrobial agents were identified. These included three ESBL genes and four genes conferring resistance to aminopenicillins (Table 4.3). Genes encoding β-lactamases that confer resistance to aminopenicillins included *bla*$_{\text{TEM-1}}$ (78.8% [74/94] isolates), *bla*$_{\text{OXA-1}}$ (21.3% [20/94] isolates), *bla*$_{\text{SCO-1}}$ (2.1% [2/94] isolates), and *bla*$_{\text{SHV-1}}$ (1.0% [1/94] isolates). 22.3%

(21/94) isolates carried at least one ESBL gene, *bla*$_{\text{CTX-M-15}}$, being the most common (Table 4.3) and was present in 95.2% (20/21) of the isolates carrying an ESBL gene. The *bla*$_{\text{CTX-M-15}}$ gene was significantly associated with resistance to cephalosporins (p<0.001) with 19/20 isolates carrying *bla*$_{\text{-CTX-M-15}}$ gene being resistant to ceftriaxone. Susceptibility profile of the remaining isolate with *bla*$_{\text{CTX-M15}}$ gene to ceftriaxone was unknown. Other ESBLs detected included *bla*$_{\text{SHV-12}}$ (one isolate) and *bla*$_{\text{OXA-10}}$ (one isolate) and both isolates had expressed phenotypic resistance to ceftriaxone. Similar to all common AMR genes identified in this collection, the *bla*$_{\text{CTX-M-15}}$ gene was present in isolates independent of phylogenetic clustering. The distribution of isolates carrying the *bla*$_{\text{CTX-M-15}}$ showed that the gene was present in isolates from all five identified phylogroups and in 11 different STs. Prevalence of *bla*$_{\text{CTXM-15}}$ by ST showed that ST391 had the highest prevalence of isolates with the *bla*$_{\text{CTX-M-15}}$ (4/4 isolates) followed by ST648 (3/3 isolates) and ST131 (21.4% [3/14] isolates) (Table 4.4). Elsewhere, the *bla*$_{\text{CTX-M-15}}$ gene is mostly identified in genomes of isolates in clade C of ST131 isolates (Petty et al., 2014). Of the three *bla*$_{\text{CTX-M-15}}$ harbouring ST131 isolates identified within the Malawian collection, two belonged to ST131 clade C, whereas one belonged to clade D, which clustered the majority of the Malawian isolates. Other ESBL genes were *bla*$_{\text{SHV-12}}$ identified in 2.1% (2/94) isolates and *bla*$_{\text{OXA-10}}$ identified in 1.0% (1/94) isolates.

Table 4.4 Characteristics of CTX-M-15 associated *E. coli* isolates

Isolate ID	Year	Source	Sequence type (ST)	Phylo group	Genome locus	Plasmid replicon
D42544	2007	Blood	1084	D	Plasmid	FIB, FII, HI2
BKQA8N	2013	Blood	131	B2	Chromosome	FIB, FII,HI2, P, Q1
C4	-	CSF	131	B2	Chromosome	Col, FIA, FIB, Q1
B28	2009	RS	131	B2	Chromosome	Col, FIA, FII, X4, Y
C30	2009	RS	167	A	Plasmid	FIA, FIB,FII, I2,Y,
C33b	-	CSF	167	A	Plasmid	FIA, FIB,FII, I2,,P, Y,
1012184	2011	Blood	361	A	Plasmid	FIB, FII
D40034	2006	Blood	391	D	Plasmid	FIB, FIA, X1
BKQ5JN	2013	Blood	391	D	Plasmid	FIB, FIA, X1,X4
D40059A	2006	Blood	391	D	Plasmid	FIA, FIB, X1
C14	-	CSF	391	D	Plasmid	A/C, FIA, FIB, FII, Q1
A1a	2009	RS	399	A	Plasmid	HI2, Y
A16	2009	Blood	448	B1	Plasmid	A/C, FIA, FIB, FII, Q1
A3b	2009	RS	617	A	Plasmid	FIA, FIB, FII
A27	2009	RS	648	F	Chromosome	FIA, FIB, FII
C15	-	CSF	648	F	Chromosome	FIA, FIB, FII
C20b	2009	RS	648	F	Chromosome	Col, FIA, FIB, FII
1016948	2011	CSF	977	B1	Plasmid	Y
1014142	2011	Blood	1163	F	Plasmid	FIA, FIB, FII, Q1,Y

RS=rectal swab

4.5.2.1 Genetic environment of bla$_{\text{CTX-M-15}}$

The blaCTX-M-15 genes were found to be located in both the chromosome and plasmid sequences (Table 4.3), but in either case their location was always downstream and adjacent to the insertion element ISEcp1 (Figure 4.10). Isolates that had the blaOXA-1 gene were more likely to have a blaCTXM-15 gene ($p<0.0001$). 18/20 isolates harbouring blaCTX-M15 also harboured blaOXA-1. In addition to the blaOXA-1 gene, a majority of ESBL isolates with blaCTX-15 also carried a number of other AMR genotypes discussed in the following sections

conferring resistance to the other commonly available antimicrobial agents in Malawi.

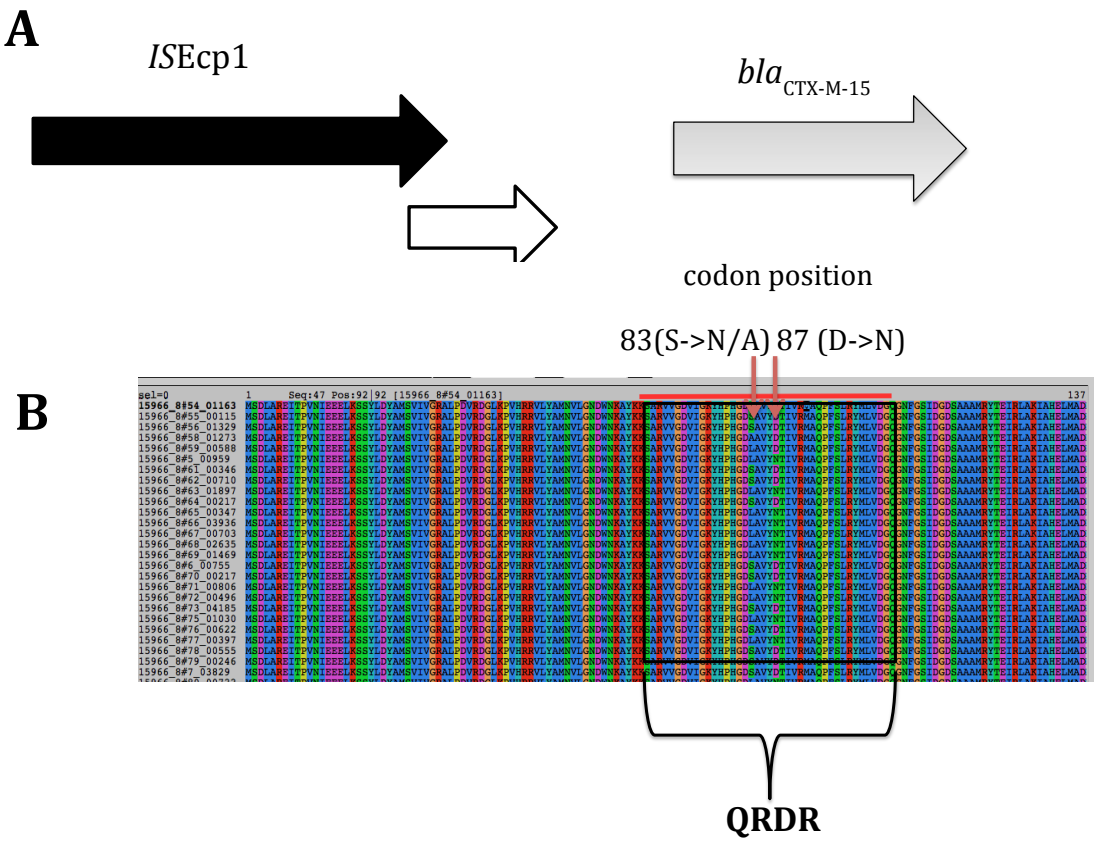

Figure 4.5 Illustration of genetic environment of main ESBL and fluoroquinolone resistance genotypes in Malawian *E. coli* isolates. (A) $bla_{\text{CTXM-15}}$ flanked by *IS*Ecp1 and truncated putative CDS upstream. (B) Amino acid alignment of GryA showing mutations in the QRDR associated with fluoroquinolone resistance.

4.5.3 Molecular determinants fluoroquinolone resistance

Protein sequence alignment of GyrA showed it was generally highly conserved. However, amino acid substitutions were observed at codon position 83 from serine

(S) to leucine (L) i.e. S83L in 23.4% (22/94) and serine (S) to alanine (A) i.e. (S83A) in 2.1% (2/94) of the isolates; and also at codon position 87, which showed a change from aspartic acid (D) to asparagine (N) i.e. D87N in 21.3% (20/94) of the isolates. Codon positions 83 and 87 of GyrA both fall within the QRDR, which ranges from codon positions 67 to 106, and mutations at these positions have previously been associated with fluoroquinolone resistance (Hooper and Jacoby, 2015a; Madurga et al., 2008). In this collection, 20 isolates had mutations at both codon positions 83 and 87 and all of them showed *in vitro* resistance to ciprofloxacin, whereas four isolates had only one codon mutation at position 83 and they were fluoroquinolone susceptible. There were amino acid substitutions outside the QRDR at positions 751 from Alanine to Serine (A751S) in 35 (37.6%) isolates. However, this mutation was not associated with ciprofloxacin resistance (p=0.102). This result suggests that to be fully resistant to ciprofloxacin, an *E. coli* isolate needed to have mutations at both codon positions 83 and 87 or at least a mutation at codon position 87, however lack of fluoroquinolone MIC data makes it impossible to determine whether the mutation at position 83 led to diminished ciprofloxacin susceptibility in these isolates. Previous studies however, suggested that a D87 mutation had a more active role in conferring resistance to fluoroquinolones than S83 mutation and that S83L was associated with higher fluoroquinolone resistance than S83A (Madurga et al., 2008). The majority of isolates that had GryA mutation at both codon position 83 and 87 also carried the ESBL gene $bla_{\text{CTX-M-15}}$. As discussed in Section 1.6.2.1, *E. coli* isolates with ESBL resistance are therefore, also likely to express fluoroquinolone resistance.

In addition to the mutations identified in the GyrA amino acid sequence, the BLAST search for acquired AMR genes against the ResFinder database, identified four isolates carrying the plasmid mediated fluoroquinolone resistance gene *qnrB*. None of the four isolates carrying the *qnrB* gene had GyrA fluoroquinolone resistance mutations and all isolates were susceptible to fluoroquinolone (Table 4.5). PMQR genes confer low-level fluoroquinolone resistance but again, the lack of MIC data here limits quantification of the effect of the *qnrB* gene on the isolates' susceptibility to fluoroquinolone.

4.5.3.1 Sequence types associated with fluoroquinolone resistance

With respect to population structure, it was noted there was a high diversity of isolates that carry genotypes associated with fluoroquinolone resistance just as there was for isolates carrying with ESBL resistance genes. Here too, isolates belonging to 11 STs (not exactly the same as those with $bla_{\text{CTXM-15}}$) carried the two mutations conferring fluoroquinolone resistance at codon positions 83 and 87 (Table 4.5). Four more isolates carrying genotypes associated with diminished fluoroquinolone susceptibility (i.e. the single mutation at position at position 83 or *qnrB* gene) belonged to four more different STs. These 15 STs spanned all the five *E. coli* phylogroups identified from this collection (Table 4.5).

Table 4.5 Characteristics of isolates with fluoroquinolone resistance genotypes*

Strain ID	ST	Phylogroup	CIP	*bla*$_{CTXM-15}$	Codon83 mutation	Codon87 mutation	*qnrB*
C12359	ST38	D	S	-	+	-	-
D48799	ST131	B2	S	-	+	-	-
A32883	ST131	B2	S	-	+	-	-
A45016	ST636	B2	S	-	+	-	-
D40034	ST391	D	R	+	+	+	-
BKQ5JN	ST391	D	R	+	+	+	-
B9070	ST10	A	R	-	+	+	-
1012184	ST361	A	R	+	+	+	-
A27	ST648	F	R	+	+	+	-
A333	ST448	B1	R	-	+	+	-
D39719	ST410	A	R	-	+	+	-
C30	ST167	A	R	+	+	+	-
C15	ST648	F	R	+	+	+	-
D40059A	ST391	D	R	+	+	+	-
C4	ST131	B2	R	+	+	+	-
A16	ST448	B1	R	+	+	+	-
C33B	ST167	A	R	+	+	+	-
C14	ST391	D	R	+	+	+	-
B28	ST131	B2	R	+	+	+	-
C20b	ST648	F	R	+	+	+	-
B3	ST617	A	R	+	+	+	-
1016948	ST977	B1	R	+	+	+	-
A3b	ST617	A	R	+	+	+	-
D45621	ST354	F	S	-	+	+	-
BKQ8AN	ST131	B2	R	+	+	+	+
9597	ST1567	D	S		-	-	+
2209	ST12	B2	S		-	-	+
D32322	ST131	B2	S		-	-	+

*R=Resistant, S=Susceptible to ciprofloxacin, + =presence of genotype, -=absence of genotype.

4.5.4 Molecular determinants of aminoglycoside resistance

Six genes known to confer resistance to aminoglycosides were identified in this dataset. These include *strA* and *strB* in 92.6% (87/94) isolates, *aac3-IIa* in 39.4% (37/94) isolates, *aac(6')-lb-cr* in 19.1% (18/94) isolates, *aph3* in 16.0% (15/94) isolates and *aadA* in 2.1% (2/94) isolates (Table 4.3). Both the *aac3-IIa* and *aac(6')-lb-cr* were associated with gentamicin and *aac(6')-lb-cr* was further associated with fluoroquinolone resistance phenotype (p<0.001). Together, *strA* and *strB* are known to confer resistance to streptomycin, which is only used in Malawi for re-treatment of tuberculosis, and there was no significant association between these genes and gentamicin resistance phenotype (p=0.318).

4.5.5 Genetic determinants of chloramphenicol resistance

Four genes known to encode chloramphenicol resistance were identified, *catA* in 64.9% (61/94) isolates, *floR* 3.2% (3/94) isolates, *catB* 2.1% (2/94) isolates and *cmlA1* 1.1% (1/94) isolates (Table 4.3). There was a significant association between presence of the *catA* gene and chloramphenicol resistance phenotype (p<0.0001). All three isolates carrying *floR* gene and the isolate with the *catB* gene, which also had the *catA* gene, had a chloramphenicol resistance phenotype whereas the isolate with *cmlA1* gene was susceptible to chloramphenicol.

In all isolates the *catA* gene, was harboured on a plasmid-borne *Tn21* transposon. In Blantyre, chloramphenicol resistance has become a marker of MDR in *Salmonella*, with almost all isolates already expressing resistance to ampicillin and cotrimoxazole (Chapter Three). The emergence of chloramphenicol resistance in *S.*

enterica in this setting was associated with insertion of the *Tn21* transposon, which carried the *catA* gene, on a plasmid of NTS strains and chromosome of *S*. Typhi strains (Wong et al., 2015; Okoro et al., 2012). The harbouring of the *catA* gene on a similar MGE in the *E. coli* strains could be suggestive of a possible exchange of AMR encoding DNA between the two species that may require further investigation.

4.5.6 Molecular determinants of cotrimoxazole resistance

Three genes known to confer cotrimoxazole resistance were identified and these included *dfrA* in 91.5% (86/94) isolates, *sul1* in 51 sul1 in 54.3% (51/94) isolates and *sul2* in 93.6% (88/04) isolates. Resistance to cotrimoxazole was strongly associated with *dfrA* gene in combination with either *sul2* or *sul1* (p<0.0001). As cotrimoxazole is a combination of trimethoprim and sulphonamide, the presence the high presence of both *dfrA* and *sul* in the *E. coli* isolates means that the majority of *E. coli* isolates have developed resistance to both compounds that make up cotrimoxazole and hence full resistance to the agent.

4.5.7 Plasmid incompatibility groups

A search against the PlasmidFinder database identified a total of 15 different plasmid replicons including FIA, FIB, FII, Col, A/C, HI1B, HI2, B/O/K/Z, I2, Q, p, p0111, R, X4 and Y to be present in the Malawian collection of *E. coli* isolates (Figure 4.9). The FIB and FIA, FII were the most common plasmid replicons across all the 94 isolates (Figure 2) but also amongst isolates which had the $bla_{\text{CTX-M-15}}$ gene (Table 4.4). One IncFIB plasmid was similar to pNDM-Mar1 (Genbank accession number

JN420336), a plasmid known to encode the carbapenem resistance gene *bla*$_{\text{NDM-1}}$, although no carbapenem resistance gene was identified in this dataset.

4.6 Discussion

The work in this chapter investigated the population structure and genetic determinants of AMR in *E. coli* isolates collected in Blantyre, Malawi, a setting from which there is limited understanding on the genetic relatedness and determinants of AMR of *E. coli*, despite it being one of the common causes of SBI. It was shown that *E. coli* in Blantyre has an expanding pan genome and high diversity in phylogenetic clustering, STs and phylogroups, consistent with the known global diversity of *E. coli* (Chaudhuri and Henderson, 2012). Furthermore, AMR in Malawi has emerged across a range of *E. coli* lineages containing a number of similar resistance genes.

The most commonly identified ESBL gene *bla*$_{\text{CTX-M-15}}$ was present in a wide range of STs across five identified phylogroups, but none of the identified STs showed dominance in CTX-M-15 production. Within SSA, high diversity of CTX-M-15 associated STs has also been reported in Tanzania and Angola (Mshana et al., 2016; Ribeiro et al., 2016) and there too, no dominant CTX-M-15 producing *E. coli* ST was identified. Globally, the *bla*$_{\text{CTX-M-15}}$ gene has been more strongly associated with a sub-lineage of ST131, *H30*-Rx (Price et al., 2013). The predominance of this ST131 sub-lineage as a CTXM-15 clone has been reported in regions across the global including Asia, Europe, North Africa, and South America (Nicolas-Chanoine et al., 2014; Ferjani et al., 2014; Chen et al., 2014; Hussain et al., 2014; Nicolas-Chanoine et

al., 2008). The results from this chapter and from other studies from the region (Mshana et al., 2016; Ribeiro et al., 2016) however, do suggest a higher diversity of CTX-M-15 STs than has been reported elsewhere, with so far, a clear dominant CTX-M-15 ST yet to be reported.

The presence of *bla*$_{\text{CTX-M-15}}$ gene in genetically heterogeneous lineages further suggests that lineage independent HGT rather than clonal expansion is the dominant mechanism for the dissemination of this gene among different strains in this population. This is different from findings in studies from outside SSA where clonal expansion ST131 CTX-M-15 sub-lineage has been identified as the main mechanism for the global dissemination of CTX-M-15 ESBL resistance (Nicolas-Chanoine et al., 2014). Dissemination of CTX-M-15 by HGT means that *E. coli* from several unrelated lineages were able to acquire ESBL genes and and spread them further. This spread may not just be limited to the *E. coli* species only but also other closely related species such as *S. enterica*.

It is not only the case that *bla*$_{\text{CTX-M-15}}$ was distributed across phylogenetically diverse isolates; fluoroquinolone resistance genotypes and all other common AMR genes including *dfrA, sul1, sul2, strA, strB, bla*$_{\text{TEM-1}}$*, catA,* and *tetB* were also widely distributed. The set of common AMR genes namely *dfrA, sul1, sul2, strA, strB, bla*$_{TEM-1}$*, catA,* and *tetB* has previously been characterised in MDR *Salmonella* circulating in Blantyre but unlike with *E. coli,* these were associated with epidemic clades (Okoro et al., 2012; Feasey et al., 2015b).

ST131 isolates in this study constitute the most clonal sequence cluster in the population structure of our collection suggesting that they were the most recent lineage to emerge or arrive in this setting. When put in the context of these global ST131 isolates, the majority of the Malawian ST131 isolates form a sub-lineage distinct from these previously identified sub-lineages. Furthermore, whilst two of the three CTX-M-15 isolates from Malawi cluster in the CTX-M-15 associated sub-lineage C, one clusters with the rest of the other Malawian isolates in sub-clade D. The CTX-M-15 associated ST131 clone is reported to have emerged early to mid 2000s (Nicolas-Chanoine et al., 2008) and has since spread in a series of clonal expansions to become the most dominant *E. coli* clone, causing a spectrum of extra-intestinal infections including urinary tract infections, bacteraemia, pneumonia, intra-abdominal and wound infections (Chen et al., 2014). In this Malawian *E. coli* collection, the earliest CTX-M-15 associated ST131 isolate was isolated in 2009 and although CTX-M-15 ST131 clone might have emerged earlier in this setting, a previous study of ESBL producing Enterobacteriaceae in 2004-2005 did not identify CTX-M-15 producing *E. coli* (Gray, 2006). It is therefore likely that CTX-M-15 associated ST131 emerged or arrived in this setting a few years after its emergence in Europe and North America.

4.7 Limitations

Isolates selection criteria aimed at obtaining as diverse a set of isolates with different AMR profiles as possible, and in enriching for diversity, the ability to estimate prevalence of different STs in Malawi *E. coli* population was lost. However,

this approach has improved the ability to describe the population structure and enabled the correct identification of the diverse lineages associated with CTX-M-15 production. Furthermore, the selection was ST blind removing the possibility of selection bias towards a particular ST.

4.8 Conclusion

The study presented in this chapter has revealed a high diversity of lineages associated with AMR including ESBL and fluoroquinolone resistance in Malawi. The results have revealed that a diverse range of STs in the Malawian *E. coli* population are MDR and ESBL producing with no dominant ESBL ST identified. The presence of $bla_{\text{CTX-M-15}}$ in a diverse range of lineages suggested that the epidemiology of ESBL producing *E. coli* in SSA is different from what has been observed globally where a $bla_{\text{CTXM-15}}$ is associated with a single clone of *E. coli* ST131. It is also possible that in Malawi, the CTX-M-15 producing ST131 *E. coli* clone emerged very recently and will expand over time to become the most prominent ESBL lineage as the case in other settings. The presence of CTX-M15 in multiple lineages further suggests HGT rather than clonal expansion, as the main mechanism for dissemination of ESBL resistance. This raises the threat of having the majority of *E. coli* lineages in this setting, despite being phylogenetically less related, to acquire ESBL resistance

Chapter Five: Population structure, genetic diversity and antimicrobial resistance of *K. pneumoniae* isolates from Malawi

5.1 Overview

K. pneumoniae is an opportunistic pathogen and a common causative agent of HA infections. It is now also being recognised for an increasing role in causing CA infection. Clones of virulent MDR *K. pneumoniae* have been increasingly reported in a number of regions across the globe. This is also the case in Malawi where *Klebsiella* spp. isolates are resistant to the majority of locally available antimicrobial agents and have the highest rates of resistance to cephalosporins and fluoroquinolones (Chapter Three), the last resort antimicrobial agents in Malawi. The molecular mechanisms of resistance and clones of *K. pneumoniae* associated with this resistance have yet to be explored in this setting.

A number of *K. pneumoniae* isolates from the MLW archive were selected for WGS, and placed in a global context using previously sequenced *K. pneumoniae* isolates from multiple locations outside SSA, in order to:

- Determine the population structure of *K. pneumoniae* isolated in Malawi;

- Place the Malawian *K. pneumoniae* isolates within the context of the global population structure of *K. pneumoniae*;
- Identify the molecular determinants of AMR and determine how such determinants relate to the *K. pneumoniae* population structure.

The analysis identified three main lineages of *K. pneumoniae* in Malawi, which correspond to those previously defined, namely lineages KpI, KpII and KpIII. The majority of the Malawian *K. pneumoniae* isolates sequenced in this study belonged to KpI lineage, however all the three lineages exhibited high genetic diversity. Further phylogenetic analysis revealed a sub-lineage of KpI, comprising of clonally related ST14 and ST15, which were a major cause of CA infections in Malawi. A large pool of AMR genes, among them multiple ESBL and *qnr* genes, was identified in the genomes of the Malawian isolates. Plasmid mediated CTX-M-15 was the dominant mechanism of ESBL resistance. In common with *E. coli* from Malawi, AMR was not restricted to a particular clade, suggesting that dissemination of AMR in the *K. pneumoniae* population in Malawi is as a result of either a combination of horizontal gene transfer and clonal expansion or horizontal gene transfer alone.

5.2 Introduction

Globally*, K. pneumoniae* is one of the leading causes of HA-infections but has additionally been reported to be a common cause of CA-infection (Podschun and Ullmann, 1998). Furthermore, *K. pneumoniae* is particularly recognised as a pathogen of global public health importance globally due to its ability to acquire and

express diverse AMR. The majority of *K. pneumoniae* reported from across the globe are MDR and ESBL producing and more recently have started to acquire carbapenemase encoding genes (Holt et al., 2015; Breurec et al., 2013; Bowers et al., 2015). ESBL and carbapenemase producing *K. pneumoniae* are a major concern as they seriously limit the treatment options of *K. pneumoniae* infection in many settings especially those in SSA where the burden of bacteria infection is high. In Chapter Three of this thesis, I found that in Malawi, *Klebsiella* spp. is one of the commonest causes of CA-BSI. Furthermore, amongst all bacterial pathogens *K. pneumoniae* had the highest rates of resistance to cephalosporins, gentamicin and fluoroquinolones.

Due to the prominence of *K. pneumoniae* as a causative agent of HA-infections and its ability to rapidly develop resistance to any available antimicrobial agent, there has recently been an increase in studies aiming at understanding the population structure and genetic background associated with the AMR capabilities of this pathogen. These recent studies have led to an improved understanding of the diversity of *K. pneumoniae* and three major phylogroups designated KpI, KpII and KpIII have been identified within the *K. pneumoniae* population (Brisse and Verhoef, 2001; Holt et al., 2015). The three lineages are now being recognised as distinct species where KpI is identified as *K. pneumoniae*, KpII as *K. quasipneumoniae* and KpIII as *K. variicola*. Some studies have also facilitated the identification of hyper-virulent and MDR *K. pneumoniae* clones such as clonal group (CG) 258, which has among its member STs, KPC producing ST258, associated with hospital epidemics in a number of settings (Bowers et al., 2015). Another important member of CG258 is

ST11, associated with both KPC and CTX-M-15 production and has also been implicated in national nosocomial epidemics in several countries such as United Kingdom, Hungary, Spain and Korea (Calbo and Garau, 2015; Oteo et al., 2016). The studies have further helped to understand the mechanisms through which AMR spreads especially amongst KpI strains. Both HGT and clonal expansions have been implicated in the dissemination of AMR among KpI strains. For example, studies by Bowers *et al.*, Holt *et al.* and Wyres *et al.* identified substantial inter-lineage exchange of genetic material through HGT (including recombination) to have facilitated spread of AMR genes across diverse lineages (Bowers et al., 2015; Holt et al., 2015; Wyres et al., 2015). Equally, clonal expansion of specific STs such as ST11, ST15 and ST258 has been identified as a major mechanism for the global dissemination of ESBL and carbapenem resistance in *K. pneumoniae* (Woodford et al., 2011; Calbo and Garau, 2015).

While there is an increasing understanding of the diversity of *K. pneumoniae* globally, few studies have included isolates from SSA and therefore, there is limited understanding on the population genomics and antimicrobial resistance of *K. pneumoniae*. Furthermore, reported studies have often focussed on HA *K. pneumoniae* such as strains in the CG258, while it is known that MDR *K. pneumoniae* especially those producing CTX-M-15, are also a common cause of CA infections (Wyres et al., 2015; Holt et al., 2015; Bowers et al., 2015). In Malawi, although *Klebsiella* spp. was identified as one of the common causes of SBI, which expresses high levels of resistance to the reserve antimicrobial agents (See Chapter Three section 3.5), no genomic characterisation of *Klebsiella* spp. isolates has ever been

undertaken. In this Chapter, I used comparative genomics and phylogenetic analysis to provide an understating of the population structure of *K. pneumoniae* in Malawi, placing it in a global context and characterising the molecular determinants of AMR in Malawian *K. pneumoniae* isolates.

5.3 Sequenced Malawian *K. pneumoniae* isolates

Following the selection criteria outlined in Chapter Two, 88 isolates identified as *K. pneumoniae* at MLW using the microbiological identification procedures described in Chapter Two were selected for WGS at WTSI. 58 (65.9%) and 14 (15.3%) of the isolates were from blood and CSF, respectively, representing invasive isolates, where as 16 (18.8%) from rectal swab representing HA carriage isolates (Figure 5.1).

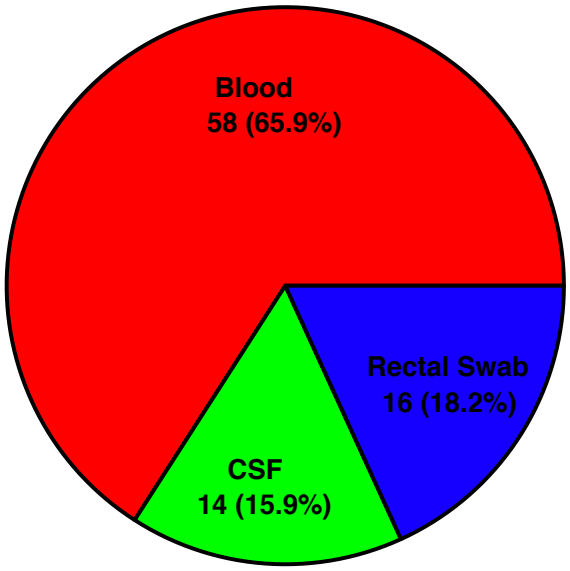

Figure 5.1 Distribution of isolates selected for WGS by clinical source of isolation.

Once the 88 isolates were sequenced, classification based on WGS match by Kraken showed that 74 genome sequences matched *K. pneumoniae,* whereas nine genome sequences matched *E. coli*, one matched *Acinetobacter baumannii*, one matched *Citrobacter rodintium*, one *S. enterica* and two *Bacillus thuringiensis* (Appendix 5). The nine genome sequences identified as *E. coli* were incorporated into the *E. coli* genome sequence dataset, whereas the others that were neither *E. coli* nor *K. pneumoniae* were excluded from further analysis.

Although *K. pneumoniae* was the top match for the remaining 74 genome sequences, the percentage identity for 3/74 isolates was very low (<25%; Appendix 3) and all three were removed from further analysis. One genome sequence, though with a relatively higher percentage identity (45%), had an excessively higher genome size (~9MB, Appendix 3) and this genome sequence too was excluded from further analysis as the genomic DNA of the isolate was likely to have been contaminated. Three genome sequences of isolates initially identified as *E. coli*, but matched *K. pneumoniae* as described in Chapter Four were incorporated into the *K. pneumoniae* dataset to give a total of 73 genome sequences of *K. pneumoniae* available for further analysis (Figure 5.2). The disagreement between the characterisation by culture and *in silico* could be explained by the same reasons as stated in Chapter Four (Section 4.4).

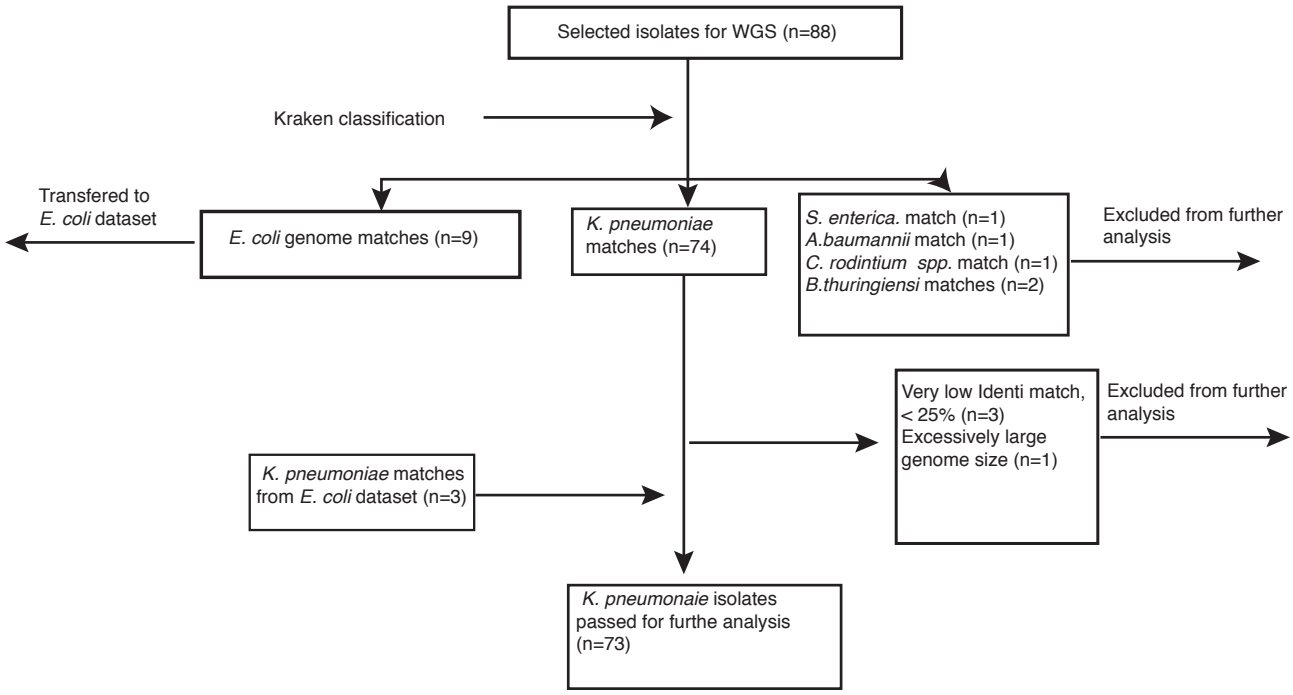

Figure 5.2 A flow-chart showing the exclusion and inclusion criteria of *K. pneumoniae* genome sequences for further analyses after quality control checks

5.4 Genetic diversity of Malawian *K. pneumoniae* isolates

5.4.1 Multi-locus sequence types and polysaccharide capsule types of Malawian *K. pneumoniae*

In silico MLST showed that the Malawian collection of *K. pneumoniae* isolates belonged to 41 different STs, although 90.2% (37/41) STs were represented by ≤ 2 isolates (Table 5.1). ST14 with 15.5% (11/71) of isolates was the most common ST, followed by ST664 (5.6% (4/71) isolates and ST15 and ST48 (4.2% (3/71) isolates each. The Simpson's diversity index of STs in this sample was 0.95, implying there is

a 95% chance that any randomly chosen isolates *of K. pneumoniae* in this setting will belong to two different STs.

Table 5.1 Distribution of STs in Malawian *K. pneumoniae* isolates

	No. of isolates(n)	$(n/N)^2$
ST14	11	0.031
ST664	4	0.0042
ST15	3	0.0023
ST48	3	0.0023
ST25	2	0.0010
ST101	2	0.0010
ST198	2	0.0010
ST268	2	0.0010
ST307	2	0.0010
ST4	1	0.00026
ST17	1	0.00026
ST20	1	0.00026
ST34	1	0.00026
ST37	1	0.00026
ST39	1	0.00026
ST45	1	0.00026
ST73	1	0.00026
ST106	1	0.00026
ST218	1	0.00026
ST231	1	0.00026
ST276	1	0.00026
ST283	1	0.00026
ST290	1	0.00026
ST297	1	0.00026
ST340	1	0.00026
ST348	1	0.00026
ST372	1	0.00026
ST383	1	0.00026
ST478	1	0.00026
ST560	1	0.00026
ST609	1	0.00026
ST874	1	0.00026
ST881	1	0.00026
ST896	1	0.00026
ST1023	1	0.00026
ST1128	1	0.00026
ST1139	1	0.00026
ST1425	1	0.00026
ST1916	1	0.00026
Simpson's index (D),		**0.054**

$$Simpson's\ diversity\ index = I - D$$

$$= 1 - 0.05$$

$$= 0.95$$

A mapping of ST identifiers to the phylogeny showed some clustering of isolates by sequence type. ST14 isolates formed the largest and most clonal sub-lineage of K-SC1 whereas other clades were mostly comprised of diverse STs (Figure 5.5). The *K. pneumoniae* genome sequences were further characterised by their polysaccharide capsule (K) types. The polysaccharide capsule is an important virulence factor of *K. pneumoniae*. K-typing of the isolates was done by identifying capsule polysaccharide sequence (*cps*) clusters using the findKcluster.py script. as seen for the MLST-derived characterisation, *K. pneumoniae* isolates had diverse K-types. In total, 24 serotypes were identified, of which K2 was the most common and was exclusively associated with ST14 isolates (Figure 5.5). The K-typing further showed that no more than one ST was linked to a serotype (Figure 5.5).

5.4.2 Pan genome and core genome of Malawian *K. pneumoniae* isolates

A preliminary pan genome analysis of the 73 remaining *K. pneumoniae* isolates produced a core genome of 109 genes. This was a too small core genome relative to other reported core genome sizes of *K. pneumoniae*, even with much larger sample sizes (Holt et al., 2015; Bialek-Davenet et al., 2014). A phylogeny was constructed on the core genome alignment obtained from the concatenation of the alignments of the 109 identified core genes. Two isolates were identified as outliers on the

resulting phylogenetic tree and were removed from further analysis. A BLAST search of contigs against the NCBI database identified the two genome sequences to be of *Klebsiella oxytoca*. The remaining 71 *K. pneumoniae* isolates, yielded an average of 5.2MBp per genome. Automatic annotation of these 71 genome sequences predicted a total gene pool of 18,521 unique protein-coding genes, which formed the pan genome. The core genome of these 71 isolates consisted of 2,820 genes (15.2% of pan genome) encoded on 2.8MB. When considered with respect to the 5.2MB average genome size of the *K. pneumoniae* isolates in this study, the core genome identified represents 53.8% of the average genome sizes of the Malawian *K. pneumoniae* isolates included in the study. This core genome also had a total of 289,173 SNP sites. The accessory genome, comprising of the remaining 15,723 genes identified in < 99.0% of the isolates comprised of genes that were uncommon with 12,357 (78.6% of accessory genome) being present in < 15.0% of the isolates and a substantial number (6,074 [49.2%]) of such genes were singletons.

The number of genes in the pan genome, core genome, set of unique genes and set of new gene per additional genome was iteratively calculated for 10 different sets of genome sequences at different sample sizes ranging between 1 and 71. The pan genome accumulation curve showed an increasing pan genome size with increasing genome sequences included in the pan genome analysis, although the rate of this increase appeared to be declining (Figure 5.3A). In comparison, the core genome size declined with increasing number of genome sequences. Similarly, number of new genes obtained from any additional genome sequences declined (Figure 5.3C

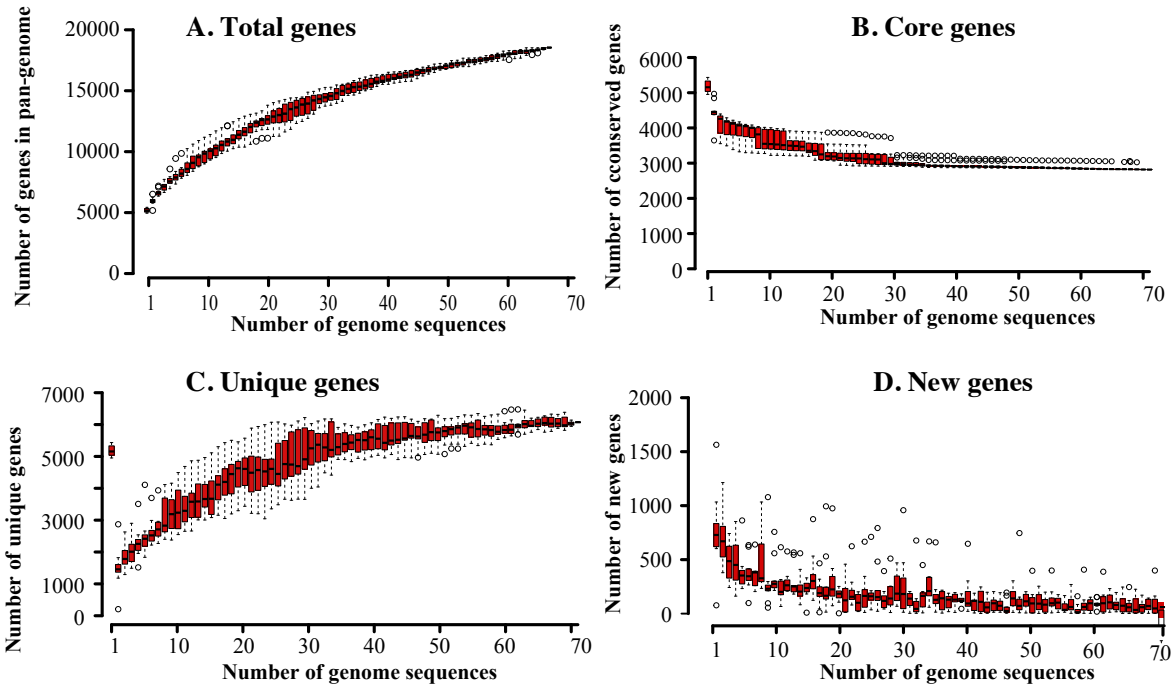

Figure 5.3 Distributions of genes in per given number of genome sequences of *K. pneumoniae* isolates from Blantyre, Malawi. (A) Total number of single copy genes identified in the pan genome of given set of Malawian *K. pneumoniae* genome sequences. (B) Number of genes conserved in at least 99% of a given set of genome sequences. (C) Number of genes unique genes i.e. identified in only one isolate as the number of *K. pneumoniae* isolates increased. (D) Number of new genes identified per additional genome sequence of *K. pneumoniae*.

5.4.3 The Malawian *K. pneumoniae* population has an open pan genome

To determine whether the pan genome curve of the Malawian *K. pneumoniae* isolates was bounded or unbounded, the relationship between the pan genome size and the number of genome sequences was subsequently modelled using the power-

law regression equation as previously described in Chapter Two. This modelling showed a significantly strong log-linear relationship ($R^2 = 0.99; p < 0.0001$) between the pan genome size and number of genome sequences. The parameters β and γ in the power-law regression equation:

$N = \beta g^{\gamma}$ (where $N = pan\ genome\ size$ and $g = number\ of\ genome\ sequences$, were estimated by the log-linear regression equation:

$$\ln(N) = \ln(\beta) + \gamma \ln(g)$$

Results of the regression analysis of the above equation yielded the following values when fitted with the data on number of genes in the pan genome:

$$\ln(\beta) = 8.43 \Rightarrow \beta = e^{8.43}$$

and

$$\gamma = 0.33$$

hence the pan genome cumulative curve (Figure 5.4) was defined by the equation.

$$N = e^{8.43} g^{0.33}$$

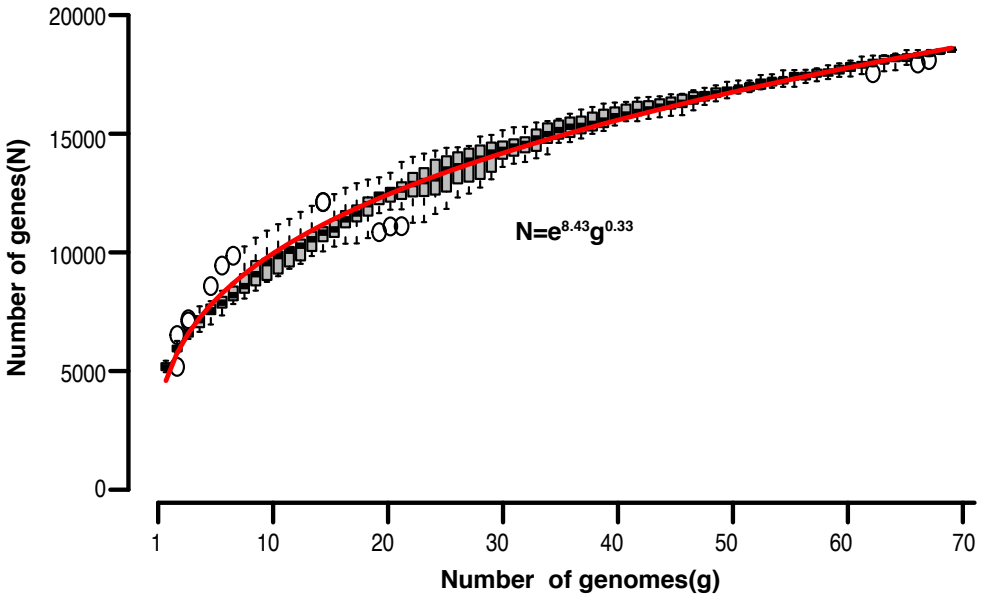

Figure 5.4 Expansion of the pan genome size with increasing number of genome sequences as modelled by the power law-regression equation $F(N) = \beta N^{\gamma}$ **illustrating an open pan genome of the *K. pneumoniae* isolates in Malawi.** By normalising the number of genes obtained in the 10 iterative samples per given set of genome sequences as described in Figure 5.3A by their mean, the values of β and γ were estimated to obtain the power-law regression equation $N = e^{8.43} g^{0.33}$ depicted by the red line-graph. The box plots black represent the distributions of number of genes in the pan genome of 10 samples of a given set of genome sequences.

Since $\gamma\ (= 0.33) > 0$, the curve defined by the equation $N = e^{8.43} g^{0.33}$ is unbounded, implying that the *K. pneumoniae* population in Blantyre has an open pan genome, i.e. inclusion of additional genome sequences in the pan genome analysis will result in further expansion of the pan genome.

5.5 Population structure of Malawian *K. pneumoniae* isolates

The population structure of the Malawian *K. pneumoniae* isolates was inferred using two approaches including a Bayesian analysis approach done by the hierBAPS module in BAPS clustering and evolutionary trees to reconstruct the phylogeny of the isolates.

5.5.1 Bayesian population structure of Malawi *K. pneumoniae*

The genome sequence reads of the Malawian *K. pneumoniae* isolates were mapped to the chromosome sequence of the *K. pneumoniae* reference strain MGH 78578 (Accession number CP000647). A Bayesian population structure analysis using the hierBAPS module in BAPS package was conducted on the alignment resulting from the mapping after Gubbins was run to remove recombination sites. BAPS clustered the isolates into three primary sequence clusters (SCs) designated K-SC1, K-SC2 and K-SC3. The majority of the isolates (66/71 [93.0%]) belonged to the cluster K-SC1, whereas clusters K-SC2 and K-SC3 had two and three isolates respectively. Calculation of pairwise core genome SNP differences showed that all the three SCs exhibited high nucleotide sequence diversity, confirming the high plasticity of the *K. pneumoniae* genomes in this population. SNP differences from the K-SC1 had a mean pairwise SNP difference of 15,020 core SNPs where as the average pairwise SNP difference for the three isolates in K-SC2 was 18,015 and the two isolates in K-SC3 differed by 18,882 SNPs.

5.5.2 Reconstructed phylogeny of Malawian *K. pneumoniae* isolates

A core genome alignment was generated by concatenating alignments of all the 2,820 core genes identified in section 5.5.1. RAxML was run on this alignment, to construct a core genome ML-phylogenetic tree. The reconstructed ML-phylogeny clustered the isolates into three major lineages that corresponded to the previously identified BAPS clusters K-SC1, KSC and K-SC3 (Figure 5.5)

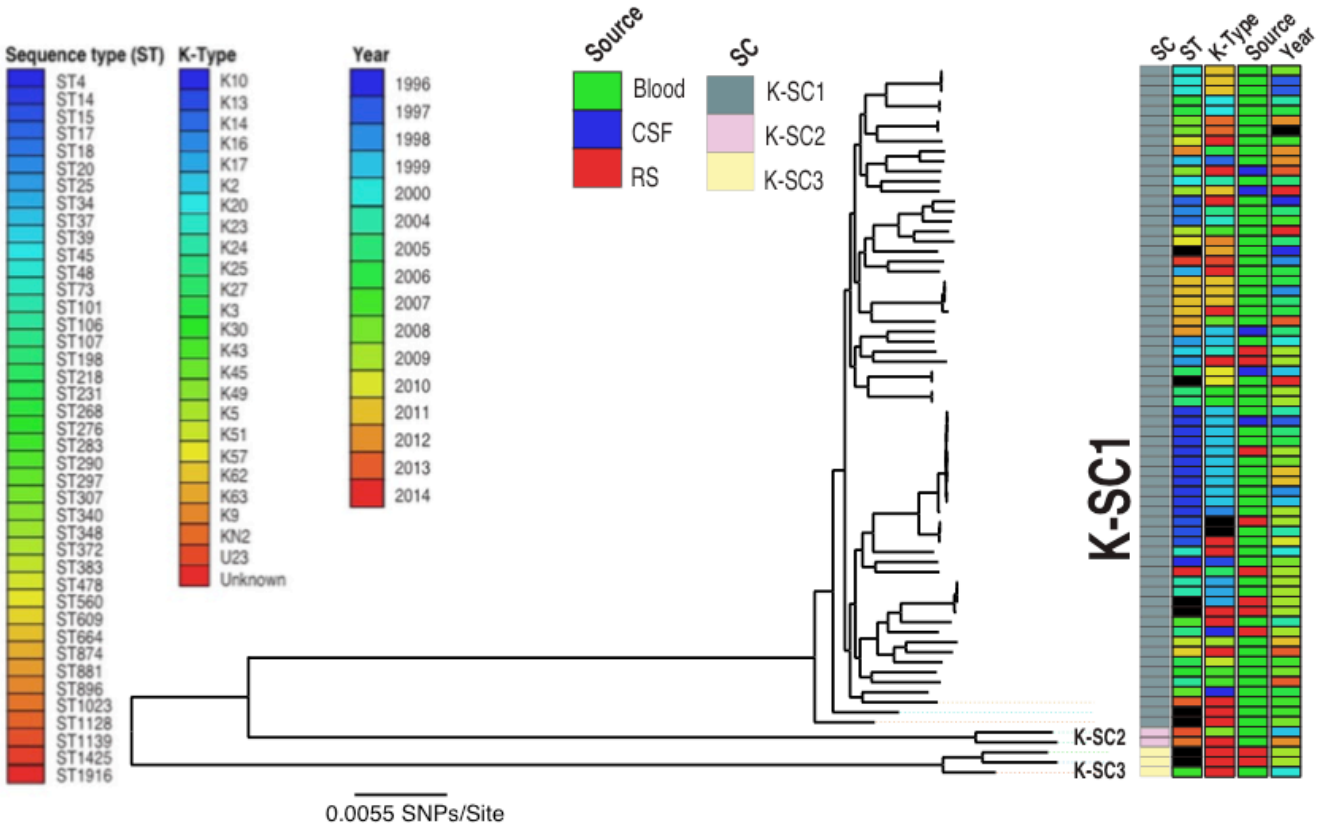

Figure 5.5 Population structure of Malawian *K. pneumoniae* isolates. The phylogeny was reconstructed using a 2.8-Mb multiple-sequence alignment with 289,173 SNPs from 2,820 core genes (present ≥ 99% of the genome sequences in single copies) and rooted at the middle of the branch separating the two most divergent sequences.

When data on year of isolation and clinical source of isolates were mapped to the phylogenetic tree, it was observed that phylogenetic clustering of isolates was independent of either the clinical source or year of isolation. More importantly it was also noted that rectal swab carriage isolates were randomly distributed across the phylogeny together with the invasive isolates, suggesting a role of carriage strains as a reservoir of pathogenic strains (Figure 5.5).

5.5.3 Malawian *K. pneumoniae* isolates in context of global *K. pneumoniae* population structure

In order to understand the population of the Malawian *K. pneumoniae* in the context of the global population structure of *K. pneumoniae*, 30 isolate genome sequences were selected from a previously published global collection where the major *K. pneumoniae* lineages (phylogroups) KpI, KpII and KpIII, were defined (Holt et al., 2015). The 30 genome sequences were selected in such a way that each of the three lineages were represented, hence the selected sample comprised of 22 genome sequences from KpI lineage, five genome sequences from KpII lineage and three genome sequences from KpIII lineage. A core genome of these genome sequences and the genome sequences of the 71 Malawian *K. pneumoniae* isolates was determined using the Roary pan genome analysis tool (Page et al., 2015). This pan genome analysis resulted in the identification of a total 28,438 genes, 53.5% larger than the pan genome size of the collection including Malawian isolates only. The core genome of the Malawian and global *K. pneumoniae* isolates consisted of 2,532, a 10.2% decline from the core genome of the collection of isolates from the Malawian

collection only- a reflection of increased diversity of the accessory genome as a result of the inclusion of the genome sequences from the global collection.

Similar to the construction of the Malawian core genome phylogenetic tree, a core genome alignment was generated from the alignments of the core genes and used to construct a phylogenetic split network using SplitsTree package (Huson, 1998). The clustering of the isolates in the phylogenetic split network revealed a correspondence between the three clusters identified in the population structure and phylogeny of the Malawian *K. pneumoniae* isolates with the previously defined global phylogroups of KpI, KpII and KpIII. The largest SC of the Malawian *K. pneumoniae* isolates, K-SC1, belonged to the lineage that is now identified as the *K. pneumoniae* proper, KpI, K-SC2 corresponded to the *K. variicola* lineage and K-SC3 corresponded to the *K. quasipneumoniae* lineage KpIII (Figure 5.6).

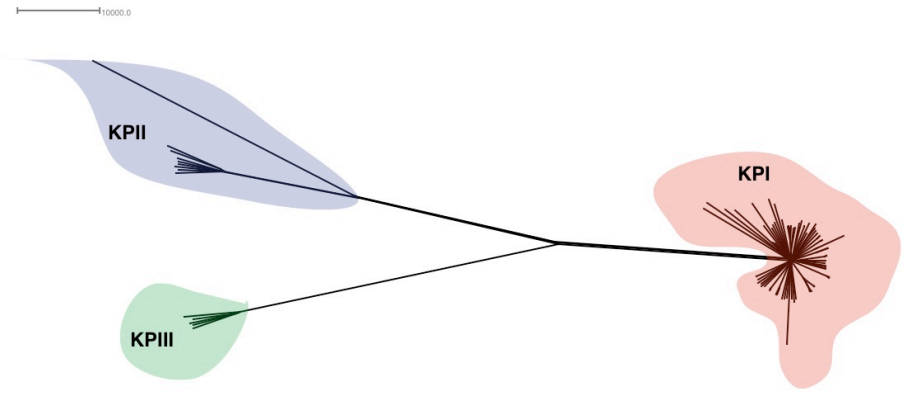

Figure 5.6 Phylogroups of Malawian *K. pneumoniae* isolates. A Split network constructed from the core genome alignment of 71 Malawian and 30 global isolates to identify phylogroups to which the Malawian *K. pneumoniae* isolates belonged.

5.5.4 Diversity of Malawian and global *K. pneumoniae* accessory genomes

The diversity of the accessory genomes of both the Malawian and global *K. pneumoniae* isolates was further investigated by principal component analysis (PCA) based on origin (global versus Malawian) and on clinical source of isolation (blood, CSF or rectal swab). PCA was applied to the accessory genes present in no more than 95% and in at least 5% of all Malawian and global *K. pneumoniae* isolates. The PCA showed three distinct clusters of isolates based on their accessory genome. These PCA clusters corresponded to the three *K. pneumoniae* lineages of KpI, KpII and KpIII previously identified using the BAPS and phylogenetic analyses as (Figure 5.7A).

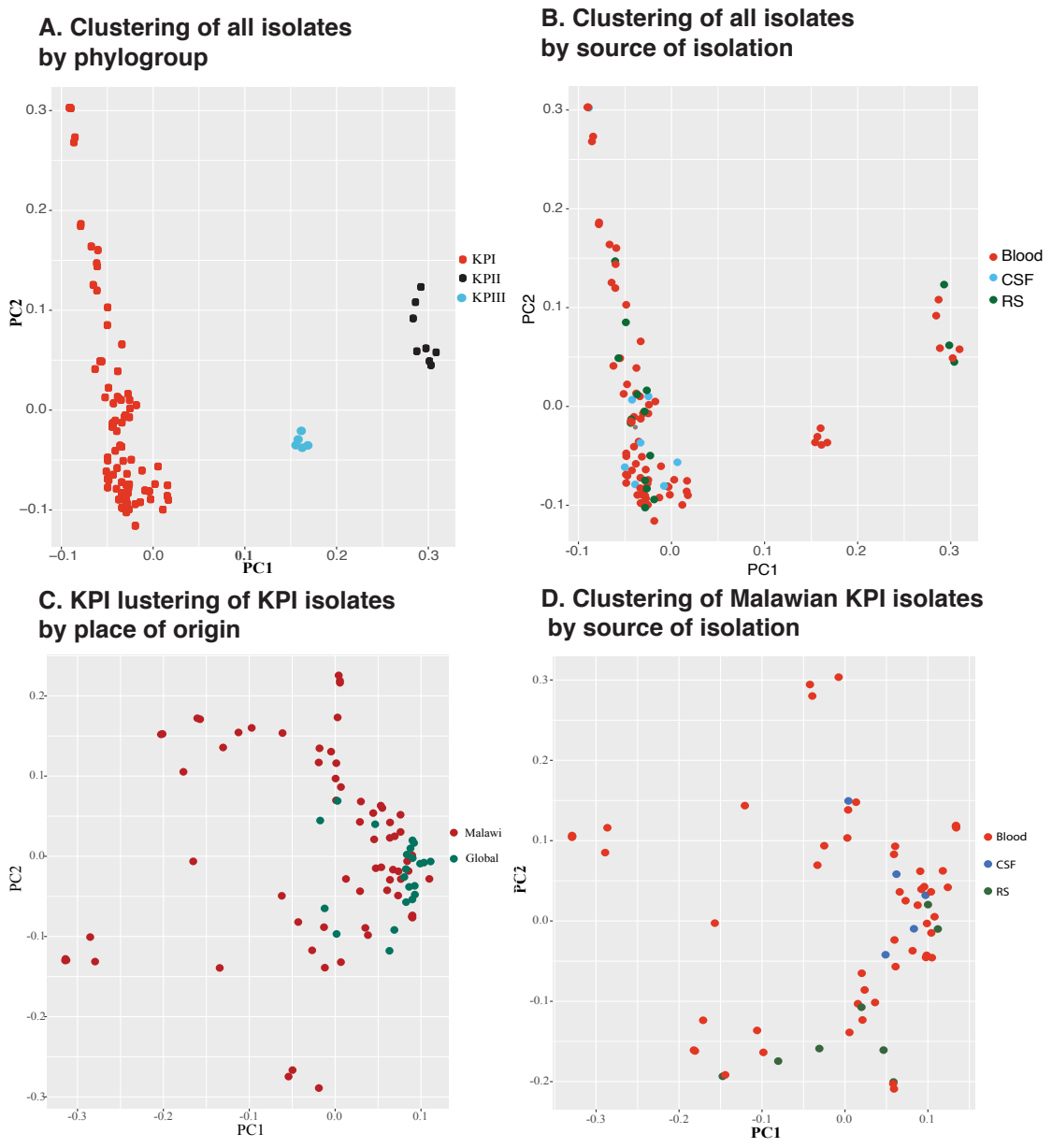

Figure 5.7 PCA Analysis of *K. pneumoniae* genomes based on accessory genes present in 5.0% - 95.0%of isolates. (A) PCA cluster of *K. pneumoniae* by accessory genes, coloured by phylogroup. (B) PCA cluster isolates by accessory genes, coloured by source of isolation. (C) PCA clustering of KpI isolates only by accessory gens and coloured by origin (D) PCA clustering of Malawian KpI isolates only, coloured by source of isolation.

Considering that the accessory genome contains genes acquired mainly through HGT (Jackson et al., 2011), the consistency between the clustering of isolates based on the core genome phylogeny and the PCA clustering of isolates based on the accessory genomes indicates that genetic differences occurring through HGT were not sufficient enough to disrupt the population structure of *K. pneumoniae* in this setting.

The PCA analysis appears to show that the accessory genome clusters KpII into rectal swab and blood isolates whereas KpIII is just a cluster of blood isolates only (Figure 5.7B). However, the low number of genome sequences available for KpII and KpIII isolates would not permit making conclusions about differences in the diversity of the accessory genome in relation to the clinical sources of samples. However, in KpI, which has a relatively large enough number of genomes, diversity of the accessory genome is independent of clinical source of isolation whether considered in Malawi only or in global context (Figure 5.7B & D). There appeared to be substantial genetic differences in the accessory genomes of the Malawian and the global collections of KpI isolates. Moreover, the Malawian isolates appear to have a more diverse accessory genome than isolates from the global collection. A comparison of the core genome of the Malawian and global KpI isolates only, revealed high gene differences between the two sets of genome sequences. Overall, the union of the core genomes of both sets of KpI isolates had 3,528 genes. The two KpI core genome sets shared 2,068 genes whereas 1,173 were specific to the Malawian collection only and 287 were specific to the global collection only (Figure 5.8).

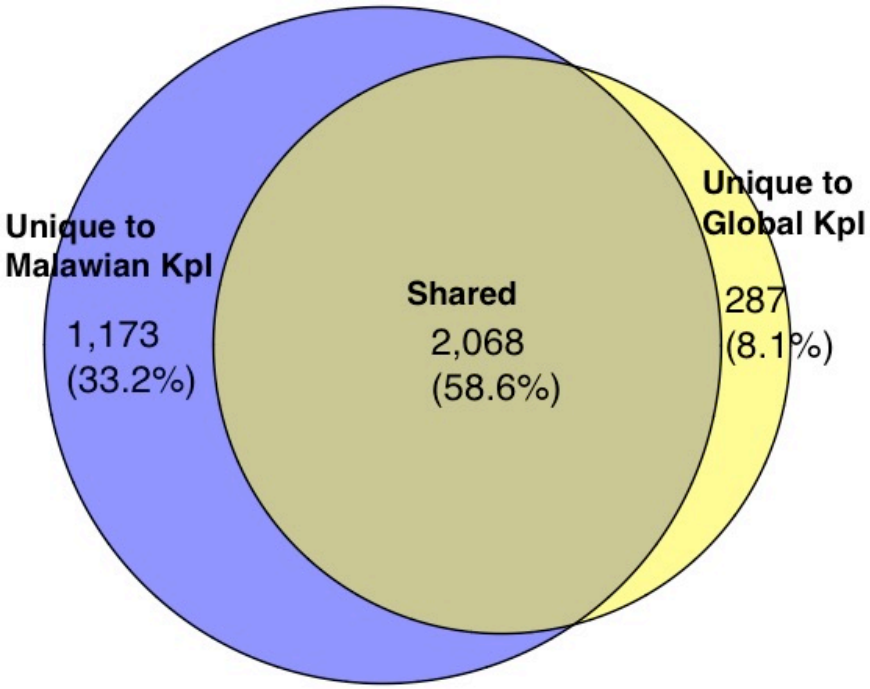

Figure 5.8 Venn diagram illustration of number of genes in the core genomes of 66 Malawian and 22 global KpI isolates.

5.5.5 Population structure of Malawian KpI isolates: Emergence and evolution of the ST14 and ST15 lineage

The K-SC1 genome sequence reads were further mapped to the genome sequence of the KpI reference strain NTUH2044 (Accession number AP006725). Gubbins was run to predict and remove recombination sites from the resulting multiple genome sequence alignment. When BAPS was applied to the recombination-free multiple sequence alignment, two sub-clusters were identified, designated KpI-SC1 and KpI-SC2. KpI-SC1 was a monophyletic clade consisting of, predominantly ST14 (11

isolates)and three other isolates belonging to ST15 (Figure 5.9). KpI-SC2 was a polyphyletic cluster containing isolates with high sequence diversity. As these genome sequences could not be assigned to a monophyletic cluster, their clustering in KpI-SC2 did not reflect common evolutionary history. Both KpI-SC1 and KpI-SC2 sub clusters consisted of both invasive and carriage rectal swab isolates.

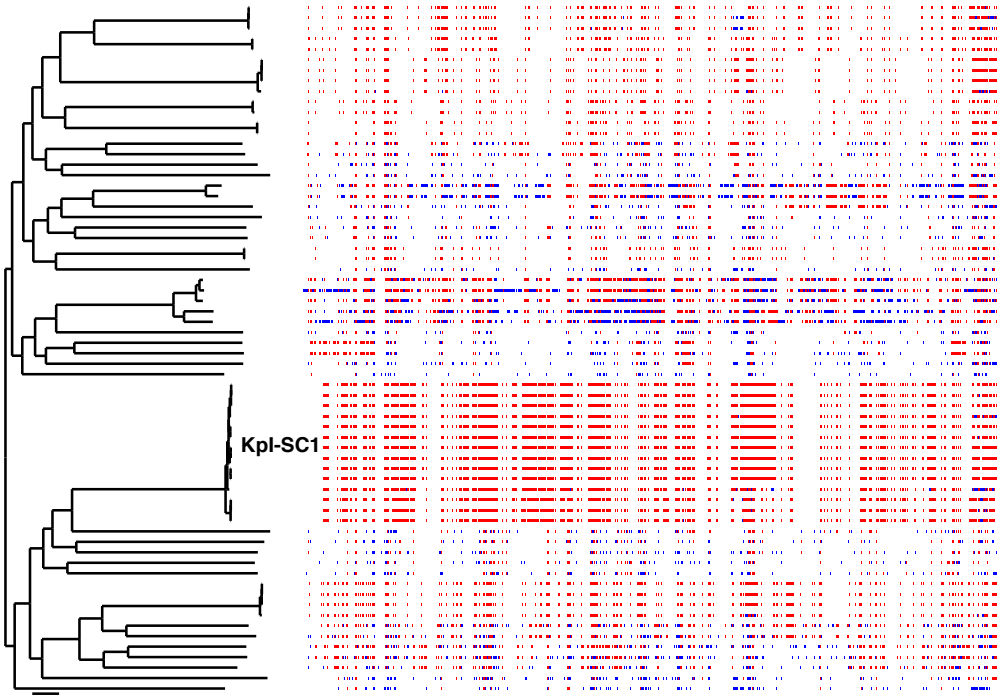

Figure 5.9 Recombination sites in genomes of Malawian KPI isolates. On the left is a ML phylogenetic tree constructed from an alignment of genome sequences (after removal of recombination sites) obtained by mapping genome read sequences of Malawian KPI isolates to the genome sequence of reference strain NTUH20144. On the right are recombination sites as predicted by Gubbins.

5.5.5.1 Recombination analysis of KpI-SC1

The recombination analysis by Gubbins revealed that K-SC1 was a highly recombinant lineage with the majority of these recombination events observed in KpI-SC1 (Figure 5.9). To elucidate the role of recombination in the evolution of both ST14 and ST15, Gubbins was further run on the genome sequence alignment of KpI-SC1 isolates only, using a sliding window of five SNPs.

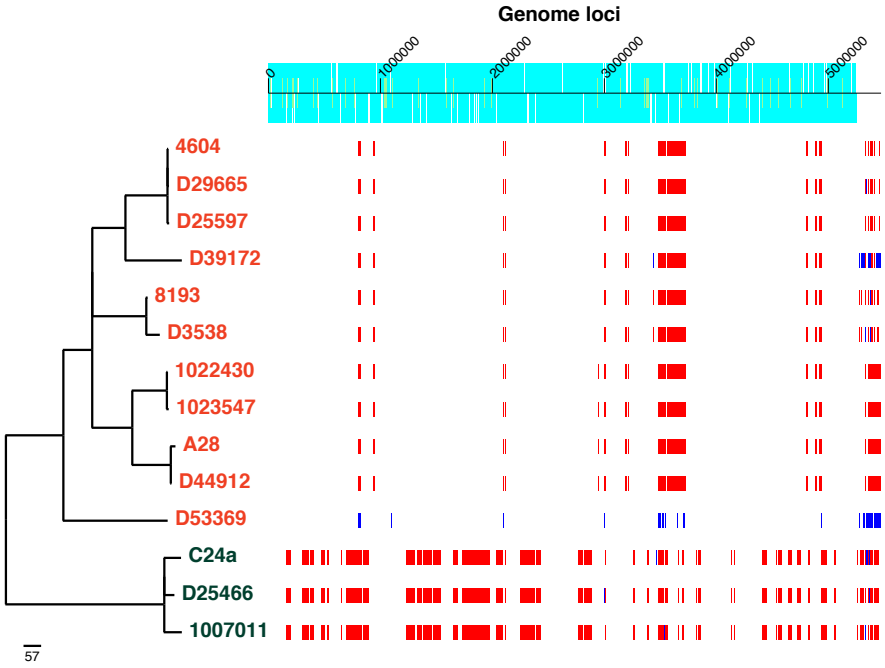

Figure 5.10 Recombination sites in genomes of isolates in KpI-SC1 sub-lineage. On the left is a ML phylogenetic tree constructed from an alignment of genome sequences (after removal of recombination sites) obtained by mapping genome read sequences of KpI-SC1 isolates to the genome sequence of the *K. pneumoniae* reference strain NTUH20144. ST14 isolates are labelled in red and ST14 isolates in green. On the right are recombination sites as predicted by Gubbins.

This recombination analysis further showed that ST15 was a more recombinant ST than ST14 (Figure 5.10). The mean recombination rate (r/m) for ST14 isolates was 1.3, where as the mean recombination rate for ST15 isolates was 4.9 Implying that ST14 isolates acquired SNPs through recombination and random mutations at an almost equal rate. ST15 isolates acquired five times more SNPs through recombination than through random mutations. This suggests that the evolution of ST15 is driven by recombination events more than random mutations unlike ST14. There were substantial disparities in frequency of recombination events in ST14 isolate (range: 0-15) while in all ST15 isolates, recombination occurred in two blocks of size > 1660kb (Table 5.2).

Table 5.2 Recombination statistics of isolates in KpI-SC1 sub-lineage

Isolate ID	ST	No. of recombination Blocks	Recombination size (kb)	Recombination sites/mutations (r/m)
D29665	ST14	1	377.4	1.5
4604	ST14	0	377.4	0
A28	ST14	0	458.8	0
1022430	ST14	0	458.8	0
8193	ST14	1	389.3	0.6
1023547	ST14	0	458.8	0
D25597	ST14	0	377.4	0
D53369	ST14	15	224.9	8.5
D44912	ST14	0	458.8	0
D3538	ST14	2	389.3	1.3
D39172	ST14	5	462.9	2.2
C24a	ST15	2	1660.1	2.28
522B	ST15	2	1630.2	1.7
1007011	ST15	2	1646.3	10.8

Annotation of the predicted recombination sites revealed that these loci encompassed a wide range of genes, transposons and insertion elements. Importantly, the analysis predicted recombination events in loci harbouring virulence factors such as the *cps* and fimbriae genes including *wzi and fim* respectively, and AMR determinants such as the *Tn3*, transposon known to contain beta-lactam resistance genes and major facilitator superfamily (MFS) that facilitate the efflux pumping of multiple antimicrobial agents (Appendix 8).

5.5.5.2 Emergence and Evolution of the ST14-ST15 lineage

KpI-SC1 being a monophyletic clade of K-SC1 implied that strains in this lineage shared a common evolutionary history. To estimate the emergence time and evolution of this lineage, BEAST analysis was run to investigate the dates of the most recent common ancestor (MRCA) and the evolutionary rates of this lineage in this setting. A genome wide SNP alignment of the isolates in this sub-clade was generated from KpI-polymorphic sites alignment, obtained after removal of predicted recombinant regions by Gubbins. A regression model of root-to-tip distance as a function of sampling time (years) was fitted on the phylogenetic tree of KpI-SC1 sub-lineage generated in section 5.6.6.1 above using TempEST software (Rambaut et al., 2016). The results of this regression analysis revealed a weak temporal signal with the coefficient of determination $R^2 = 0.36$ and estimated KpI-SC1 to have emerged from its MRCA estimated in 1833 (Figure 5.11).

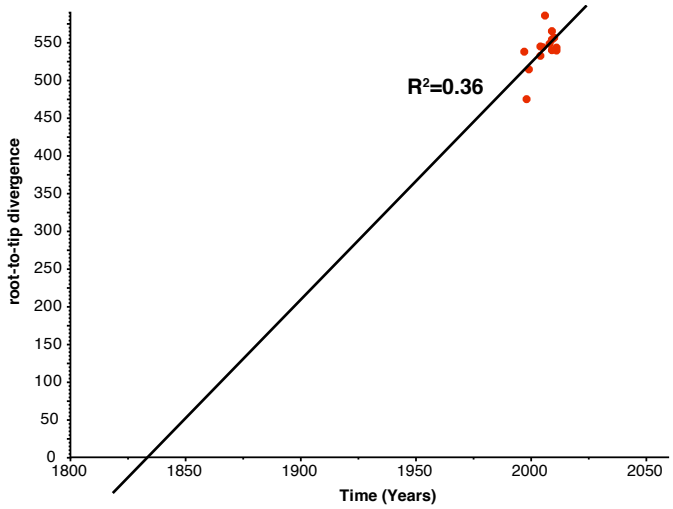

Figure 5.11 Estimation of root-to-tip divergence over sampling time of Malawian *K-pneumoniae* isolates belonging to the KpI sub-cluster KpI-SC1.

The signal was however still considered sufficient for estimation of substitution rates and evolutionary time of this lineage in Malawi. Previously, studies have estimated the evolutionary rate and time of *K. pneumoniae* based on weaker signal than the one identified here. For example, a study by Duchêne *et al.* estimating evolutionary rates of a number of bacterial species used much weaker signal of $R^2 = 0.25$ to estimate the evolutionary rate of one of the three clades of *K. pneumoniae* CC258 (Duchene et al., 2016).

The evolutionary rate and time were estimated using the Bayesian Markov Chain Monte Carlo (MCMC) method in BEASTv1.7.5 with a chain length of 100,000,000, sampling every 20,000 steps. As with the reconstruction of maximum likelihood phylogenies, the GTR substitution model was used with a Gamma rate of heterogeneity. Further model options included uncorrelated lognormal molecular

clock model and constant-size coalescent and Bayesian Skyline model.

Posterior node heights were estimated from the maximum clade credibility tree and showed that the KpI-SC1 diverged from its MRCA ~155 (95% HPD [14.1,445.4]) years to the most recent isolate in the collection (Figure 5.12). The analysis further suggested that ST15 diverged from ST14 44 95% HPD [9.0, 117.7) years ago (Figure 5.12). Overall substitution rate was estimated at $\sim 3.92 \times 10^{-7}$, 95% HPD [1.2×10^{-10}, 1.3×10^{-6}]. Whilst the regression analysis of divergence over sampling time by TempEST above showed a weak signal for temporal structure and the BEAST estimates are characterised by wide credible intervals, the estimates here, especially for ST15, agree with previous estimates from a study by Moradigaravand, *et al* (Moradigaravand et al., 2017), which estimated the emergence of ST15 in Europe in 1965.

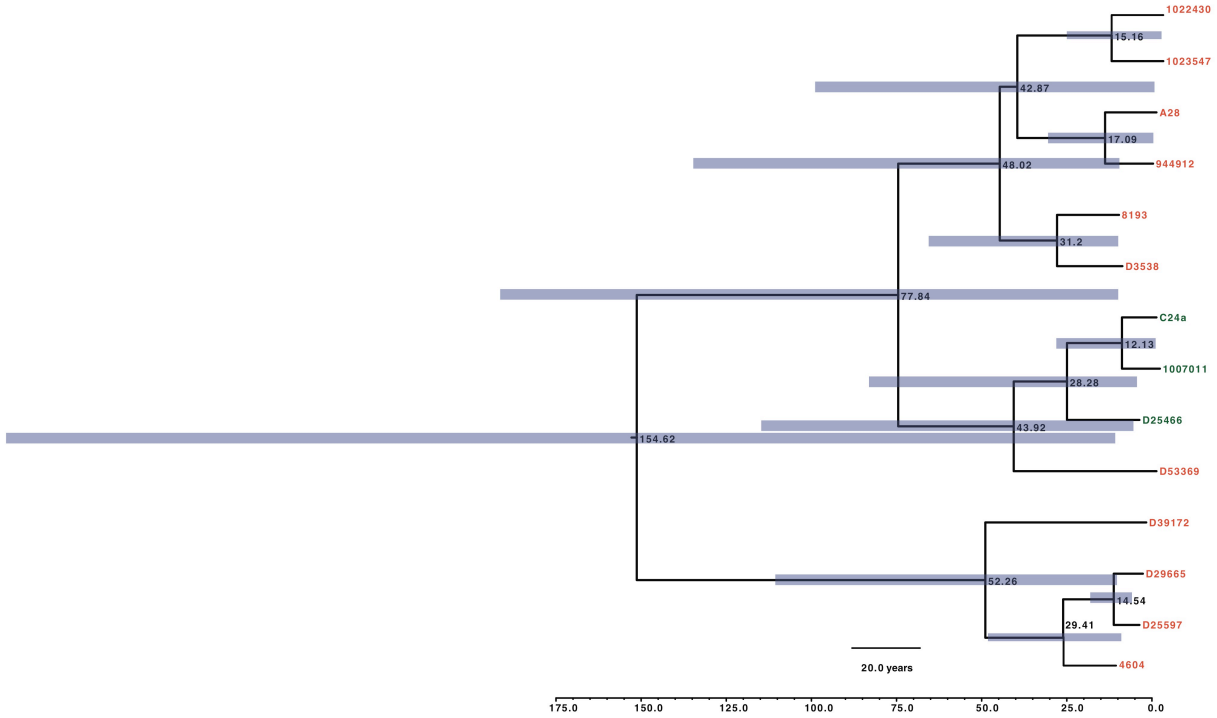

Figure 5.12 Dated Bayesian phylogenetic tree of KpI-SC1 comprising of ST14 (with red labels) and ST15 (with green labels) isolates. The values on nodes denote divergence times in years whilst the blue bars across nodes represent the 95%HPD intervals of the divergence times.

5.6 Genetic determinants of antimicrobial resistance

5.6.1 Acquired AMR genes in genomes of Malawian *K. pneumoniae*

BLAST search against the ResFinder database of acquired AMR genes identified a total of 43 unique AMR genes. A full list of all of the identified AMR genes in the genomes of the *K. pneumoniae* isolates are shown in Appendix 6. The identified genes are known to confer resistance to a wide range of antimicrobial agents such as beta-lactams (excluding carbapenems and cephamycins), chloramphenicol, aminoglycosides, fluoroquinolones, sulphonamides, trimethoprim, macrolides and tetracyclines (Table 5.3) and here, analysis of the AMR genes and phenotype data show a general concordance between the genes and the AMR phenotypes (Figure 5.13).

Table 5.0.3 Distribution of AMR genes present in at least five isolates. (A list of all 43 AMR genes identified from the genomes of the *K. pneumoniae* isolates is presented in Appendix 7)

AMR gene	Description	Class of antibiotics	Prevalence	
			n	%
fosA	metalloglutathione transferase	Fosmycin	69	97.2%
oqxA	Oqx Efflux pump gene	Fluoroquinolones	67	94.4%
oqxB	Efflux pump gene	Fluoroquinolones	66	93.0%
Sul2	Sulfonamide-resistant dihydropteroate synthase	Sulphanomides/cotrimoxazole	56	78.9%
dfrA	Dihydrofolate reductase	Methaxazole	58	81.7%
aac(6')	Acetyltransferase	Aminoglycoside, Fluoroquinolone	53	74.6%
$bla_{\text{TEM-1}}$	Beta-lactamase	Aminopenicillins	53	74.6%
catA	acetyltransferase	Chloramphenicol	53	74.6%
strB	Streptomycin phosphotransferase	Aminoglycosides	46	64.8%
strA	Streptomycin phosphotransferase	Aminoglycosides	46	64.8%
$bla_{\text{CTX-M-15}}$	ESBL	Beta-lactams	28	39.4%
Sul1	Sulfonamide-resistant dihydropteroate synthase	Sulphanomides/cotrimoxazole	25	35.2%
$bla_{\text{SHV-1}}$	Beta-lactamase	Beta-lactams	22	31.0%
tetD	Tetracycline efflux gene	Tetracyclines	21	29.6%
$bla_{\text{SHV-11}}$	Beta-lactamase	Beta-lactams	17	23.9%
mphA	Macrolide phosphotransferase	Chloramphenicol	16	22.5%
tetA	Tetracycline efflux gene	Tetracyclines	12	16.9%
$bla_{\text{OXA-1}}$	Beta-lactamase gene	Beta-lactams	11	15.5%
aadA2	Aminoglycoside adenylyltransferase	Aminoglycosides	11	15.5%
$bla_{\text{SHV-28}}$	ESBL	Beta-lactams	10	14.1%
arr	ADP-ribosylation catalysing enzyme gene	Rifampin	7	9.9%
alph3	Aminoglycoside phosphotransferase	Aminoglycosides	6	8.5%
qnrB	PMQR gene	Fluoroquinolone	6	8.5%
cmlA1	MFS transporter/ chloramphenicol efflux gene	Chloramphenicol	5	7.0%
floR	Transmembrane segments efflux gene	Chloramphenicol/Florfenicol	5	7.0%
$bla_{\text{OXA-10}}$	ESBL	Beta-lactam	5	7.0%
tetB	Tetracycline efflux gene	Tetracycline	5	7.0%

5.6.2 Distribution of acquired AMR genes across the phylogeny of the Malawian *K. pneumoniae*

When the data on the presence or absence of AMR genes was mapped to the phylogeny of the Malawian *K. pneumoniae*, it was observed that the majority of the genomes harboured at least 10 AMR genes. Furthermore, as was the case with Malawian *E. coli* (Chapter Four), MDR was independent of phylogenetic clustering (Figure 5.13). Three genes namely *fosA*, *oqxA* and *oqxB* were present in almost all the genomes of the Malawian *K. pneumoniae*, whereas the rest were randomly distributed. The independence of acquisition of AMR genes and phylogenetic clustering suggests that the prominence of HGT as a main mechanism through which AMR is disseminating among *K. pneumoniae* in Malawi.

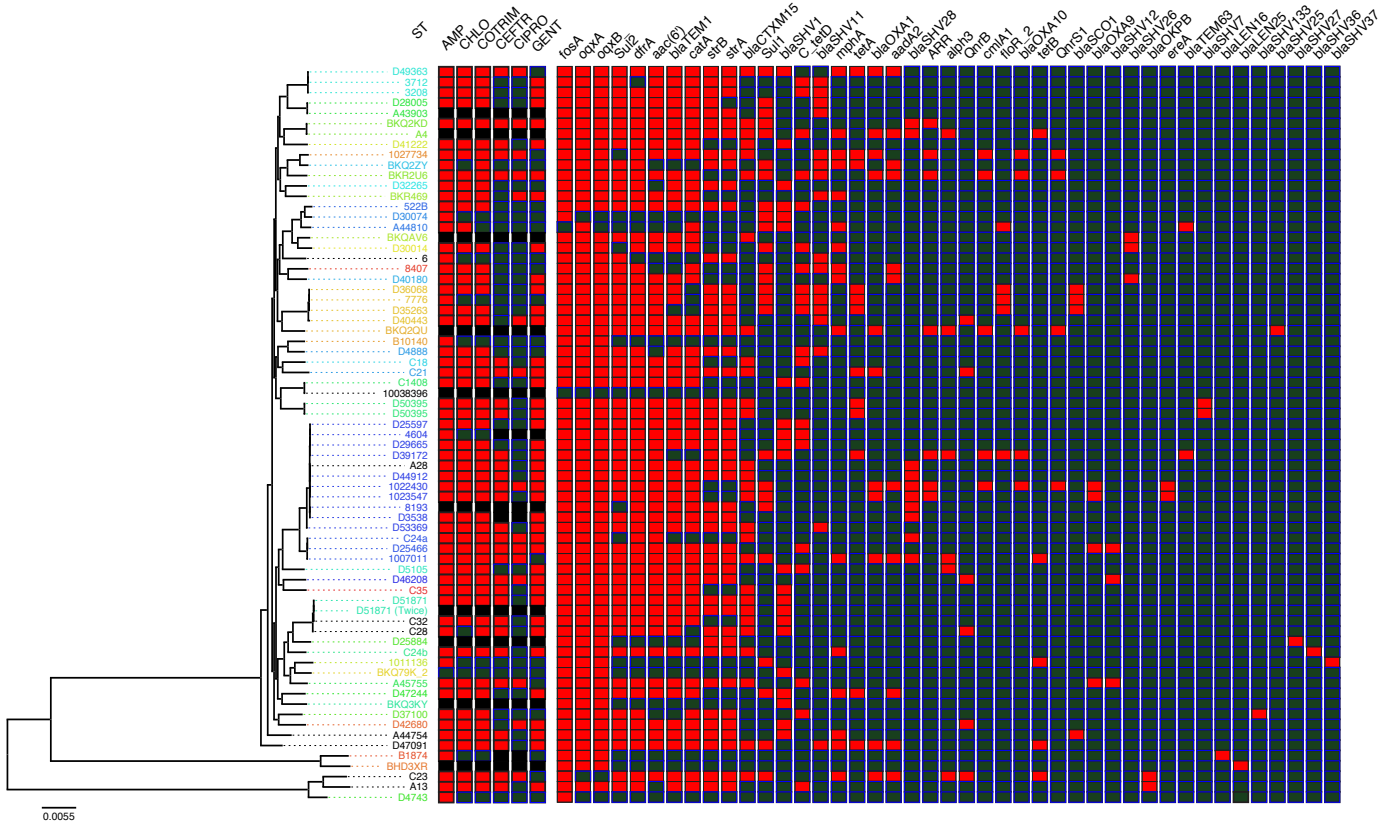

Figure 5.13 Distribution of AMR phenotypic profiles and acquired AMR genes across the phylogeny of Malawian *K. pneumoniae* isolates. The first six columns at the termini of the phylogeny represent AMR phenotypes for ampicillin, chloramphenicol, cotrimoxazole, ceftriaxone, ciprofloxacin gentamicin; red=resistant, blue=sensitive). Following the phenotypes are columns indicating presence (red) or absence (blue) of AMR genes. .

Distribution of number of genes per genome by lineage shows that KpI and KpII isolates from Malawi had relatively similar distributions of number of AMR genes per genome with both distributions characterised by substantial variations between isolates (Figure 5.14). The number of AMR genes amongst Malawian KpI isolates ranged from 0 to 19 genes per genome with average of 12 genes per genome where as amongst KpII isolates, the number of genes per genome ranged from 1 to 17 with a slightly lower mean of 10 AMR genes per genome than KpI isolates. In comparison, KpIII isolates had a substantially smaller repertoire of AMR genes when compared to either KpI or KpII isolates (Figure 5.14). Both isolates in KpIII lineage had four AMR genes, three of which (*fosA*, *oqxA* and *oqxB*) were common to both isolates where as for the fourth, each isolate had a different variant of a bla_{LEN} beta-lactamase gene. The genome of the isolate B1874 harboured $bla_{\text{LEN-16}}$ where as the genome of the isolate BHD3XR harboured $bla_{\text{LEN-25}}$ (Figure 5.13). There were no differences in terms of number of AMR genes per genome based on clinical source, but when outliers were excluded, isolates from CSF exhibited higher variations in number of genes per genome than isolates from either blood or rectal swabs (Figure 5.14B). The number of AMR genes amongst isolates from blood ranged from 0 to 19 genes per genome, 3-18 genes per genome amongst CSF isolates and 7-19 genes per genome amongst rectal swab isolates. Both blood and CSF isolates both had a mean of ~11 AMR genes per genome, where as carriage rectal swab isolates had a mean of ~12 AMR genes per genome. Whilst no significant differences of mean number of genes per genome were observed (R^2 = 0.02; p=0.2847), there was significant increase in the maximum number of AMR genes harboured per genome over time

(Figure 5.14). These results do suggest *that K. pneumoniae* and *K. quasipneumoniae* have a more expanded AMR profile than *K. variicola* strains. Furthermore, carriage isolates are a reservoir of AMR genes facilitating further spread in the community whilst the *K. pneumoniae* resistome is expanding over time.

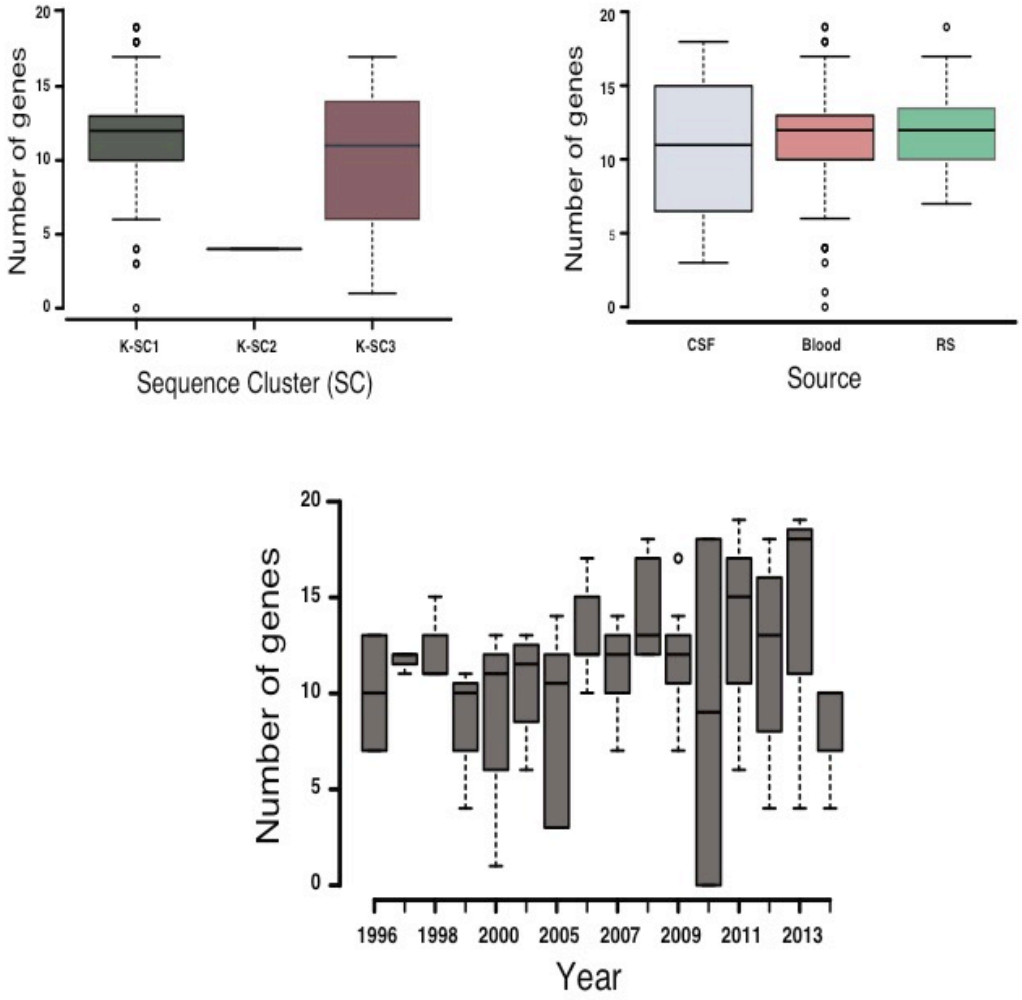

Figure 5.14 Distribution of number of AMR genes per genome (A) Distribution of AMR genes per genome by sequence cluster (B) clinical source of isolation and (C) year of isolation.

5.6.3 Molecular determinants of β-Lactam resistance

Aminopenicillins and ceftriaxone are the most commonly used β-lactam antimicrobial agents in Malawi, although *K. pneumoniae* is normally considered to be intrinsically resistant to the former due to the presence of chromosomally encoded $bla_{\text{SHV-1}}$ gene in almost all *K. pneumoniae* genomes. The analysis of the Malawian *K. pneumoniae* genomes in this thesis has revealed the presence of a high number of β-lactamase and ESBL genes, conferring resistance to both amino penicillins and cephalosporins. β-lactamase (including ESBL) resistance genes were the most common class of AMR genes identified in the genomes of the Malawian *K. pneumoniae* isolates. Twenty unique β-lactamase genes (46.5% of all AMR genes) were identified and among them were 14 distinct ESBL genes. Genes encoding resistance to the aminopecillins only included $bla_{\text{TEM-1}}$, $bla_{\text{SHV-1}}$, $bla_{\text{OXA-1}}$, $bla_{\text{SCO-1}}$, $bla_{\text{SHV-11}}$ and bla_{OKPB} (Table 5.3). Against the notion that $bla_{\text{SHV-1}}$ is core to *K. pneumoniae*, here it is noted that the $bla_{\text{SHV-1}}$ was only present in 22/71(31.0%) genome sequences (Table 5.3). $bla_{\text{CTX-M-15}}$ was the predominant ESBL gene identified in 28/71 (39.4%) isolates followed by $bla_{\text{SHV-28}}$ (10/71 [14.1%] isolates) and $bla_{\text{SHV-9}}$ and $bla_{\text{SHV-10}}$ (5/71 [7.0%] isolates) each. Others included $bla_{\text{SHV-12}}$ and $bla_{\text{SHV-26}}$ each in 3/71 4.2%] isolates, $bla_{\text{SHV-7}}$ and $bla_{\text{TEM-63}}$ each in 2/71 (2.8%) isolates, and $bla_{\text{LEN-16}}$, $bla_{\text{LEN-25}}$, $bla_{\text{SHV-133}}$, $bla_{\text{SHV-25}}$, $bla_{\text{SHV-27}}$ and $bla_{\text{SHV-36}}$ and $bla_{\text{SHV-37}}$ each present in 1/71 (1.4%) isolates (Table 5.3).

The genetic environment of $bla_{\text{CTX-M15}}$ was similar to that of *E. coli* described in Chapter Four (Section 4.6.2.1) and raising the possibility of interspecies exchange of

the DNA encoding the gene. However unlike *E. coli*, all the $bla_{\text{CTXM-15}}$ gene here was harboured on a plasmid sequence only. Again just as with *E. coli* the $bla_{\text{CTXM-15}}$ and other ESBL genes were present in isolates that were randomly distributed across the *K. pneumoniae* phylogeny (Figure 5.15) pointing to HGT being a mechanism for dissemination of ESBL resistance across distantly related *K. pneumoniae* lineages.

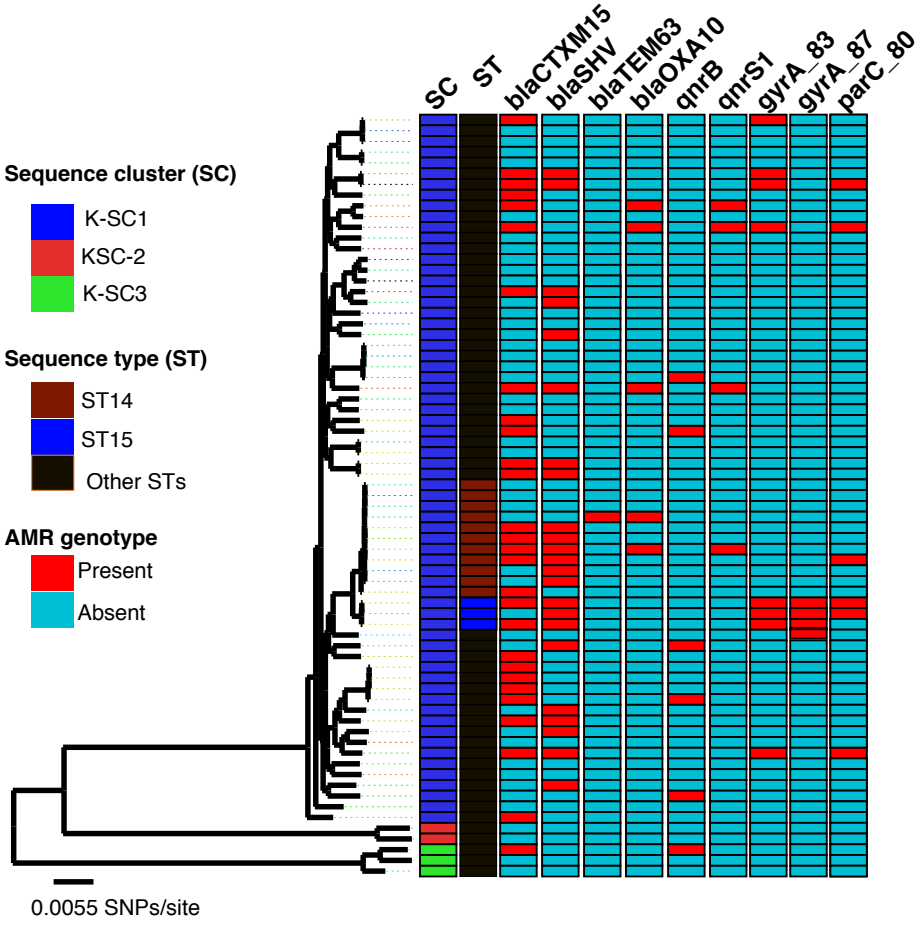

Figure 5.15 Distribution of ESBL and fluoroquinolone resistance genotypes across the phylogeny of the Malawian *K. pneumoniae* isolates.

5.6.4 Molecular determinants of fluoroquinolone resistance

Acquired AMR genes that have previously been associated with low-level fluoroquinolone resistance identified in this collection included *oqxA* (67/71 isolates), *oqxB* (66/71 [93.0%] isolates), *qnrB* (6/71 [8.5%] isolates), *qnrS* (4/71 [5.6%] isolates) and *aac(6')-lb-cr* gene. Isolates with *qnrB* or *qnrS* genes were significantly associated with resistance to ciprofloxacin (p<0.0001). Five of the six isolates with *qnrB* gene were resistant to ciprofloxacin (one isolate was susceptible) where as three of the four *qnrS1* isolates were all resistant to ciprofloxacin. Despite previously being linked to conferring fluoroquinolone resistance through efflux pumping (Chopra and Galande, 2011), *oqxA* and *oqxB* were not significantly associated with ciprofloxacin resistance in this collection (p=0.558), but it is important to note that MICs were not available to evaluate DCS.

Analysis of the amino acid sequence alignments of GyrA identified codon mutations S83I (4 isolates), S83F (3 isolates) and S80Y (1 isolate) and were all significantly associated with the ciprofloxacin resistance phenotype (p<0.0001). The three isolates with the S80F codon mutation also had a D87A amino acid substitution. One isolate with one amino acid substitution D87N was sensitive to ciprofloxacin. Other studies have also linked mutations in the QRDR of the Par*C* protein sequences to fluoroquinolone resistance especially mutations in codon positions 80 and 84. In this collection, mutations were identified at codon position 80, S80I (six isolates). Five of the six isolates with ParC mutation S80I also had either S83F and D87A or S83I only and all were resistant to ciprofloxacin. One isolate with Par*C* S80I, which

did not have a coding mutation in the QRDR of GyrA and did not have the ciprofloxacin resistance phenotype suggesting that the GryA mutation were of more importance in the expression fluoroquinolone resistance. None of the 11 isolates with either the *qnrB* or *qnrS1* genes had a GyrA or ParC QRDR (Figure 5.15) mutation hence the reduced susceptibility to ciprofloxacin resistance phenotype as described could solely be attributed the *qnrB* and *qnrS1* genes. Whilst the rest of the fluoroquinolone resistance genotypes were not lineage specific, codon mutation D87A was linked to ST15 and were all ciprofloxacin resistance placing this ST as a major fluoroquinolone resistant clone requiring further investigation in this setting.

5.6.5 Molecular determinants of aminoglycoside resistance

Five genes known to confer aminoglycoside resistance were identified. The most common was *aac(6')-lb-cr* which was present in 53/71 (74.6%) isolates. *strA* and *strB* were the second most common aminoglycoside resistance genes as they were detected in 46/71 (64.8%) of the isolates. Other less common aminoglycoside resistance genes included *aadA* and 11 (15.5%) isolates, *alph3* in six (8.5%) six isolates (Table 5.3). The only aminoglycoside resistance phenotype data available were for gentamicin and 42/60 (67.7%) isolates tested were resistant to gentamicin. 40/42 (95.2%) gentamicin resistant isolates had the *aac(6')-lb-cr* gene (p<0.0001). This result indicated *aac(6')-lb-cr* is the main gene encoding gentamicin resistance.

5.6.6 Molecular determinants of chloramphenicol resistance

The chloramphenicol resistance associated acetlytransferase (*catA*) gene was identified in 53/71 (74.6%) isolates and was significantly associated with chloramphenicol resistance phenotype in this collection (p<0.0001). Chloramphenicol resistance profile data were available for 63/71 (88.7%) of the isolates and (52/63) (82.5%) isolates had shown phenotypic resistance to chloramphenicol. 47/52 (90.4%) of the chloramphenicol resistant isolates harboured the *catA* gene. Other genes associated with chloramphenicol resistance identified in this collection included *cmlA* and *floR* genes each in 5/71 (7.0%) isolates. Four of the five isolates with the *cmlA1* gene were resistant to chloramphenicol but three of those isolates had the *catA* gene as well. Similarly, four of the five *floR* isolates were resistant to chloramphenicol but here only one isolate had the *catA* gene too.

5.6.7 Molecular determinants of cotrimoxazole resistance

Cotrimoxazole is one of the most widely used antimicrobial agents in Malawi but as described in Chapter Three, a majority of the major bacterial pathogens in Malawi have developed widespread resistance to this agent. In the collection that was selected for sequencing and passed the quality control for further analysis, 88.7% (63/71) *K. pneumoniae* isolates were tested by the disc diffusion method for resistance to cotrimoxazole and 84.1% (53/63) of the tested isolates were cotrimoxazole resistant. A search for genes known to confer resistance to cotrimoxazole revealed the presence of the *dfrA* gene in 81.7% (58/71) isolates, the

sul1 gene in 35.2% (25/71) isolates and the *sul2* gene in 78.9% (56/71) isolates (Table 5.3). Presence of *sul1* gene or *sul2* gene or both was associated with presence of the *dfrA* gene (p<0.0001) i.e. isolates with the *dfrA* gene also had either had one of *sul1* or *sul2* genes or both. All the three genes were significantly associated with resistance to cotrimoxazole (p= 0.002).

5.6.8 Plasmid incompatibility types associated with AMR

Using a BLAST search against the PlasmidFinder database, 12 plasmid replicons representing plasmid incompatibility groups were identified from the genome sequences of our *K. pneumoniae* collection. The identified plasmid replicons included incCol (six isolates), IncA/C2 (five isolates), IncFIA (15 isolates), IncFIB (54 isolates), IncFII (54 isolates), IncHI1 (17 isolates), incHI2 (one isolate), IncN (2 isolates) IncR (37 isolates) and IncQ1 (eight isolates), IncP1 (two isolates) and incX1 (1 isolate). The majority of the AMR genes were mostly associated with IncFIB and IncFII (Figure 5.16). The incFII and incFIB replicons were identified in exactly the same isolates and were associated with exactly the same genes. This suggests either that that the incFII and incFIB replicons were harboured by one plasmid or that the two replicons represent two different plasmids, which coexist to provide stability to each other.

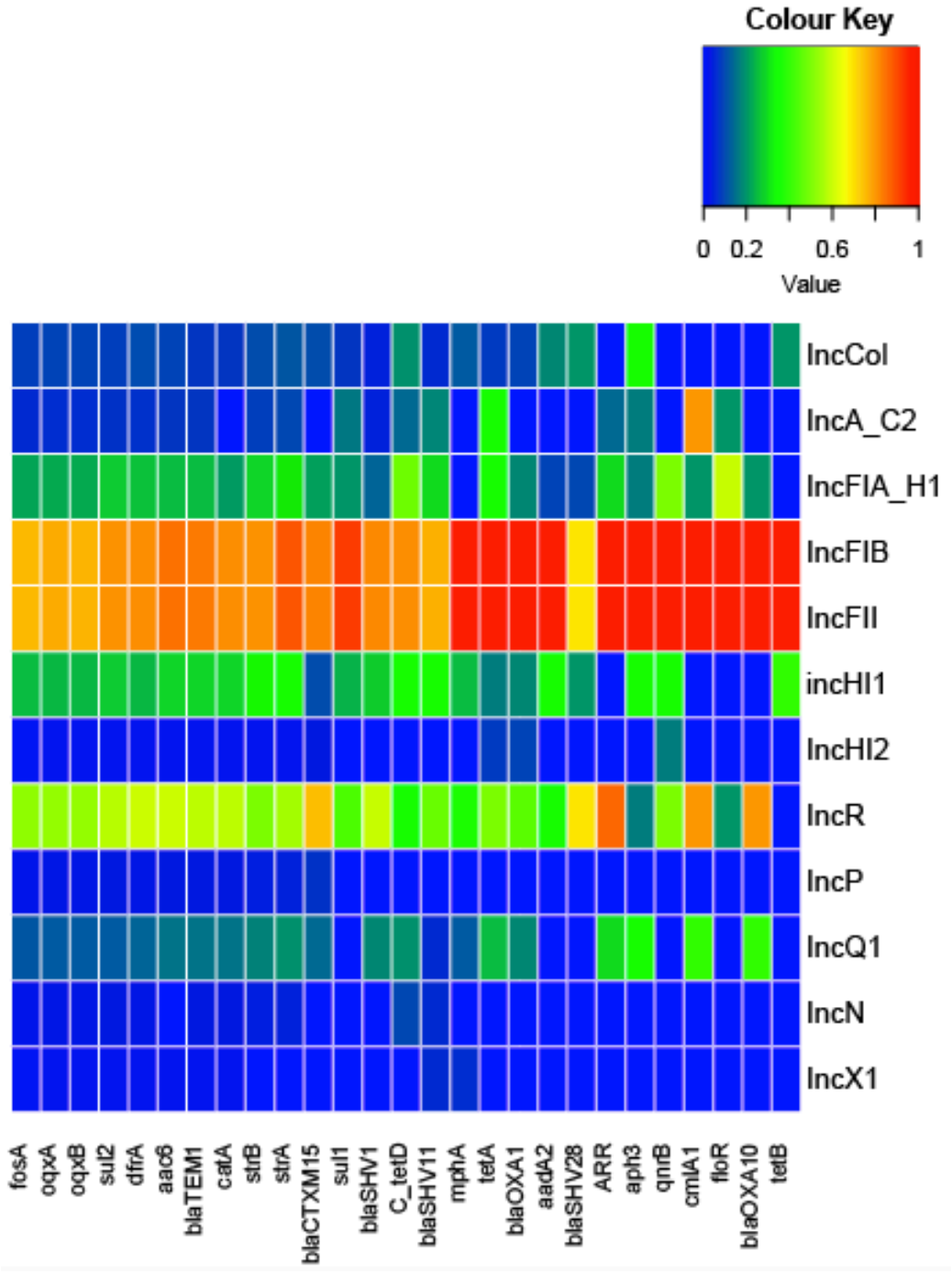

Figure 5.16 A heatmap illustration of the association between plasmid Inc-types and acquired AMR genes present in ≥ 5 genomes. Association values were measured by number of isolates plasmid isolates to number of isolates with AMR genes

5.7 Discussion

K. pneumoniae is a pathogen of global importance due to its association with extensive AMR and nosocomial infection, however little is known in the context of SSA. This is despite the fact that in SSA, *K. pneumoniae* has been identified as a common cause of both HA and CA infections and is frequently MDR (Breurec et al., 2013; Mshana et al., 2013a). In Chapter Three, *K. pneumoniae* was identified as a common cause of CA-BSI among patients presenting to QECH and worryingly, the majority of the isolates were becoming resistant to all the locally available antimicrobial agents. The genetic context of these phenotypic trends has not previously been described. In order to expand the knowledge and understanding of the population genomics and mechanisms of AMR of *K. pneumoniae* in SSA, and Malawi in particular, this chapter investigated the population structure and genomic determinants of AMR of *K. pneumoniae* isolates from Malawi both independently as well as in the context of other global isolates.

The findings of this chapter indicate the presence of a diverse *K. pneumoniae* population in Malawi, comprising of both rare or distantly related STs and a few well known STs that have been implicated in human infection globally, including SSA. Consistent with the global population structure of *K. pneumoniae*, three major phylogroups have been identified: KpI, KpII and KpIII, where KpI is the dominant phylogroup, and to lesser degree, KpII and KpIII. All three lineages, exhibited high genetic diversity, characterised by an open pan genome with a huge accessory genome and a very high core genome SNP density. The dominant KpI lineage

comprised of a large number of unrelated STs, which, with the exception of clonally related ST14 and ST15, formed a polyphyletic clade. A comparison of the Malawian KpI and global KpI pan genomes indicated that the Malawian KpI isolates had a more diverse accessory genome than the global isolates, which could suggest increased adaptability traits by the Malawian KpI strains as compared to the global strains.

KpIII has been proposed as *Klebsiella variicola*, a species very closely related to *K. pneumoniae* and methods to distinguish the two are very limited, such that *K. variicola* strains are in most cases identified as *K. pneumoniae* (Garza-Ramos et al., 2016). Previously, *rpoB* gene analysis or a multiplex-PCR method based on amplification of unique genes identified by genomic comparison of the two species has been used, but recently phylogenetic analysis was able to cluster the two species into separate phylogroup clusters (Garza-Ramos et al., 2016; Holt et al., 2015). Here *K. variicola* isolates was identified from *K. pneumoniae* isolates using WGS and phylogenetic analysis, highlighting the important role that WGS can play in the characterisation of pathogens in any setting.

Most *K. pneumoniae* studies globally and in other regions outside SSA have investigated with HA *K. pneumoniae* (Bowers et al., 2015). These studies have led to the identification of key epidemic clones, particularly of the CG258, and have noted them to be globally distributed. Here all invasive isolates were presumed CA and the analysis has revealed that a lineage comprising mostly of ST14 strains and a few ST15 strains is a major cause of CA- *K. pneumoniae* infection in this setting. The only

member of the CG258 identified in this *K. pneumoniae* collection is ST340, which had one isolate. Within SSA, the predominance of ST14 among *K. pneumoniae* population is not unique to Malawi. ST14 was identified as the most common *K. pneumoniae* ST causing paediatric HA-infection and ESBL producing in Tanzania (Mshana et al., 2013a). Whilst there have been reports of ST14 in other settings outside SSA, such reports have not identified ST14 as a predominant ST. The dominance of ST14 in several SSA countries does suggest that this *K. pneumoniae* ST has become endemic to this region and it will be important to investigate its transmission in future studies.

Both ST14 and ST15 have been associated with HA-infection elsewhere within and outside SSA (Hu et al., 2013; Mshana et al., 2013a; Breurec et al., 2013). Here, ST14 and ST15 were identified as the commonest STs isolates of CA *K. pneumoniae* In addition, sub-lineage KpI-SC1, which clusters the ST14 and ST15, also contains the hospital carriage rectal swab isolates. The identification of these international hospital STs among isolates associated with CA infection and the their phylogenetic mixing with HA carriage isolates did raise the possibility that the distinction between HA and CA strains in Blantyre is not straight forward. This could be due to ill health and poor water and sanitation hygiene (WASH) infrastructure in Blantyre and high disease burden resulting in a large proportion of the population frequently visiting QECH and getting colonised by the hospital strains, which later circulate in the community. The identification of ST14 and ST15 sub-lineage as a common cause of CA-infection here further demonstrates that this sub-lineage represents an internationally transmitted clone.

ST14 and ST15 have previously been shown to belong to different but clonally related CGs CG14 and CG15 respectively (Bialek-Davenet et al., 2014). Whilst the ML phylogeny illustrates the distinctiveness of these STs, BEAST analysis has shown evolutionary overlap of the two STs, probably owing to interlineage genetic exchanges through mechanisms such as recombination, which is evidently common between the two STs. The results here however, suggested that ST15 diverged from ST14. The BEAST analysis in this thesis has shown that this ST diverged from the ST14 lineage approximately 49 years ago. The predominance and clonal relationship between ST14 and ST15 has also been shown in another study comprising of *K. pneumoniae* isolates from five African countries and Vietnam (Breurec et al., 2013). Another study from the UK, which had also included secondary whole genome sequence data from other parts of Europe and Asia estimated that the ST15 might have emerged in Asia and Europe an estimated 48 and 37 years ago respectively (Moradigaravand et al., 2017). Based on previous estimates on the emergence of ST15 and the findings of this thesis, it is highly possible that the ST14 and ST15 diverged from their common ancestor in SSA and later spread to other parts of the world such as in Europe and Asia, where they have become some of the successful and dominant *K. pneumoniae* STs associated with AMR.

Hypervirulent K. *pneumoniae* strains causing CA- infections have been shown to predominantly belong to the capsular serotypes K1 and K2, but appear to be different from MDR clones. In defining hypervirulent and MDR *K. pneumoniae* clones, Bialek-Davenet et al (Bialek-Davenet et al., 2014), defined CC14 and CC15,

the clonal groups containing ST14 and ST15, as MDR CCs but not hypervirulent, although there were few ST14 isolates that were of serotype K2. Polysaccharide capsular typing of the Malawian *K. pneumoniae* genome sequences identified K2 as the most common serotype and was exclusively associated with ST14 (Figure 5.5). It has further been shown that, consistent with the previous reports, all ST14 isolates from Malawi are MDR and a majority are ESBL producing.

Globally *K. pneumoniae* is notorious for having high level resistance to a wide range of antimicrobial agents and in our collection, we have identified a large pool of acquired AMR genes, including ESBL genes such as $bla_{\text{CTXM-15}}$, bla_{SHV12}, bla_{SHV10}, $bla_{\text{OXA-10}}$ as well as mutations and acquired genes conferring high and low level fluoroquinolone resistance respectively. As was the case with *E. coli* (Chapter Four), $bla_{\text{CTX-M-15}}$ is the dominant ESBL gene in this collection of *K. pneumoniae* isolates. Elsewhere $bla_{\text{CTX-M-15}}$ has been detected in *E. coli* more commonly than *K. pneumoniae* isolates however, here the results suggest CTX-M15 is more prevalent in *K. pneumoniae* isolates (39.4%) than the *E. coli* isolates (21.3%) (p=0.0089). Possession of $bla_{\text{CTXM-15}}$ is mostly identified with strains causing CA infection, whereas nosocomial ESBL *K. pneumoniae* are more typically associated with possession of the bla_{SHV} and bla_{TEM} ESBL genes. The predominance of CTX-M-15 amongst the different ESBL variants identified here perhaps reflects that the *K. pneumoniae* isolates being investigated in this study are from patients with CA-infection.

Spread of MDR and indeed ESBL production in *K. pneumoniae* has been associated with the expansion of a limited number of *K. pneumoniae* epidemic clones, although the role of HGT is appreciated (Woodford et al., 2011). In such scenarios, there are AMR patterns that are lineage specific. The findings of this chapter showed that MDR or ESBL production is not restricted to the genetic background of the isolates, as no particular AMR profiles have been identified with specific lineages or STs. The strong association between the majority of the AMR genes and a limited number of plasmid replicons including the incFII, incFIB and R plasmid replicons does suggest that plasmids have a key role in harbouring and dissemination of AMR genes. This is further evidenced by the fact that even the distribution of the common plasmid replicons associated with the AMR genes is independent of the phylogenetic clustering of the isolates or their STs. In particular, the results have identified recombination as one of the drivers of the spread of AMR and virulence factors such *cps* and *fim* genes.

Although genes encoding carbapenem resistance were not identified in this collection of *K. pneumoniae,* the findings of this chapter indicate that the Malawian population of *K. pneumoniae* certainly has the potential to acquire carbapenem resistance encoding genes. The dominant ST in this collection, ST14, has acquired New Delhi Metallo-beta-lactamase one gene ($bla_{\text{NDM-1}}$) in a number of settings and indeed the first *K. pneumoniae* isolate harbouring $bla_{\text{NDM-1}}$ was an ST14 (Yong et al., 2009). Furthermore, ST14 carrying various carbapenem resistance genes have been identified in SSA countries with travel links and bacterial disease transmission history with Malawi, such as Kenya and South Africa. In Kenya, a study investigating

seven MDR *K. pneumoniae* isolates as part of surveillance of AMR mechanisms and infection control audits, found seven isolates that harboured the $bla_{\text{NDM-1}}$ gene and all the seven isolates were ST14 (Poirel et al., 2011). In South Africa ST14 isolates carrying $bla_{\text{OXA-48}}$ have also been identified (Jacobson et al., 2015). A number of isolates (including ST14) from the Malawian collection of *K. pneumoniae* have plasmids with a high sequence similarity to the pNDM-1 plasmids. These plasmids would provide the genetic environment necessary for the acquisition, persistence and dissemination of $bla_{\text{NDM-1}}$ genes, should the evolutionary selection pressure for their emergence, through dysregulated carbapenem use, be brought to bear on this population.

5.8 Limitations

As was the case in Chapter Four, the criteria for selection of isolates aimed to obtain the most diverse set of isolates as possible, with different AMR profiles. By enriching for diversity, the ability to estimate prevalence of different STs in the *K. pneumoniae* population was lost. However, as with the *E. coli* isolates in Chapter Four, the selective criteria did improve the description of the population structure and enabled the correct identification of the diverse lineages associated with ESBL production. Furthermore, the selection was "ST-blind" removing the possibility of selection bias towards a particular ST.

5.9 Conclusion

K. pneumoniae in Malawi are very diverse with a population structure characterised by globally defined phylogroups. However, the Malawian KpI has been shown to

have a more diverse accessory genome than other settings, a reflection of the heterogeneous environment from which the isolates were sourced. A diverse range of genotypes have been associated with ESBL and fluoroquinolone resistance but $bla_{\text{CTX-M15}}$ genes and mutations in the QRDR respectively, are the major genotypes. Despite the huge diversity of the *K. pneumoniae* in this setting and acquisition of MDR including ESBL genotypes by isolates across all lineages, a lineage of internationally transmitted ST14 and ST15 isolates has been identified to be a predominant cause of CA *K. pneumoniae* infection in this setting. In particular, ST14, which was associated with NDM-1 production elsewhere, has shown to be a high risk ST characterised by both MDR genes and hyper virulence factors. Whilst carbapenems would be an effective alternative antimicrobial agent for treatment of MDR *K. pneumoniae*, there is high potential of *K. pneumoniae* strains in this setting to easily acquire and spread carbapenem resistance with genetic platform for harboring and dissemination of the $bla_{\text{NDM-1}}$ already in existence.

Chapter Six: A mathematical model estimating incidence of *E. coli BSI* in a Blantyre population with heterogeneous susceptibility

6.1 Overview

This chapter presents preliminary work on mathematical modelling of the dynamics of BSI in Blantyre. The aim of the work is to estimate the force of DRI and evaluate the various factors fuelling the spread of DRI. This preliminary work has laid a foundation for a model that will achieve this goal by establishing that there is heterogeneity in susceptibility to BSI. The chapter further shows how this heterogeneity can be accounted for through a model that estimates the incidence of *E. coli* BSI over time and derives the formula for calculating the force of infection.

6.2 Introduction

Mathematical models have been used to demonstrate how interplay of various factors influences spread of infection including DRI. Estimation of parameters through mathematical models enables the evaluation of the contributions of various factors or interventions associated with an infection and permits one to make predictions about the spreading patterns of the infection in a population (Grohn et al., 2017). In the past, mathematical models assumed homogeneous mixing in the

population to lessen the complexity of the models if heterogeneity were taken into account, or due to lack of data (Keeling, 2008). In the case of susceptibility to infection this means that such models assume individuals in a population are equally susceptible to infection. However, this assumption is seldom true. In the case of susceptibility to BSI, factors that drive BSI such as malaria, HIV, and malnutrition are associated with particular subsets of individuals within a population. Furthermore, individuals may naturally be born with differential susceptibility to infection as a result of other inherent factors such as prematurity or genetic differences between individuals (Chapman and Hill, 2012).

In the past twenty years, Blantyre has experienced a high burden of BSI (Chapter Three). Attempts have been made to evaluate the impact of factors such as HIV/AIDS, malaria, malnutrition and climate have had on some of the BSI epidemics experienced using mathematical and statistical models (Feasey et al., 2015a; Pitzer et al., 2015). These models too, did not consider the possibility of susceptibility heterogeneity within the population. This chapter focuses on determining whether there is underlying heterogeneity in susceptibility to BSI in Blantyre, using *E. coli* BSI as a case study dataset. It then shows how heterogeneity in susceptibility to BSI can be accounted for in a model that estimates and predicts trends in incidence of BSI.

6.3 Population with homogeneous susceptibility to infection

Incidence of BSI was modelled under the assumption that individuals in the population are born equally susceptible to infection. Susceptible Individuals in

compartment *S* get infected at a rate λ, the infected individuals in compartment *I* recover (after receiving treatment) at rate γ. Individuals who recover become immune hence were not susceptible for re-infection.

The infection dynamics with respect to age were described by the following set of differential equations:

$$\frac{dS}{dA} = -\lambda S$$

$$\frac{dI}{dA} = \lambda S - \gamma I$$

where A is age, and the initial conditions are $S(0) = 1$, $I(0) = 0$, indicating that all individuals are uninfected at birth.

6.4 Population with discrete groups with different susceptibility levels

The population was assumed to have discrete groups with different levels of susceptibility to infection. Here, two groups S_2 and S_2were considered with susceptibility factors $\delta_1$and δ_2, respectively. Under these assumptions the infection dynamics are now defined by the following set of differential equations:

$$\frac{dS_1}{dA} = -\delta_1 \lambda S_1$$

$$\frac{dS_2}{dA} = -\delta_2 \lambda S_2$$

$$\frac{dI_2}{dA} = \delta_1 S_1 + \delta_2 \lambda S_2$$

With initial conditions

$S_1(A) = \rho_1, S_2(A) = \rho_2,\ I_1(A) = I_2(A) = 0$, such that

$S_1(0) = \rho_1, S_2(0) = \rho_2, I_1(0) = 0, I_2(0) = 0$ such that $\rho_1 + \rho_2 = 1$, and $\rho_1\delta_1 + \rho_2\delta_2 = 1$ where $\rho_1$and ρ_2 are the proportions of individuals born with risk factors $\delta_1$and δ_2, respectively.

For both models, it was assumed that *death rate* = *birth rate*. Recovery rate was given by $\gamma = \frac{1}{duration}$. Based on available clinical information, patients with bacterial BSI take approximately 7 days (0.25 *months*) to recover hence $\gamma = (\frac{1}{4})^{-1} = 4$, (when age is in months) or $\gamma = (\frac{1}{52})^{-1} = 52$ (when age is in years).

Both models were first fit on a subpopulation that was < 5 years old where incidence was stratified into monthly age groups. The best fitting model of the two was then extended to fit to the incidence data for all age groups and then over time.

Observed annual and age-stratified incidence of *E. coli* was as calculated from the surveillance data as described in Chapter Two.

6.5 A heterogeneous susceptibility model provides a better fit to observed incidence

The two models were implemented in MATLAB. When fitted against the age stratified incidence of E. coli BSI in children under five years of age (in months), the model under the assumption of a heterogeneous susceptibility population provided a better fit to incidence data than the model that assumed homogeneous susceptibility (Figure 6.1).

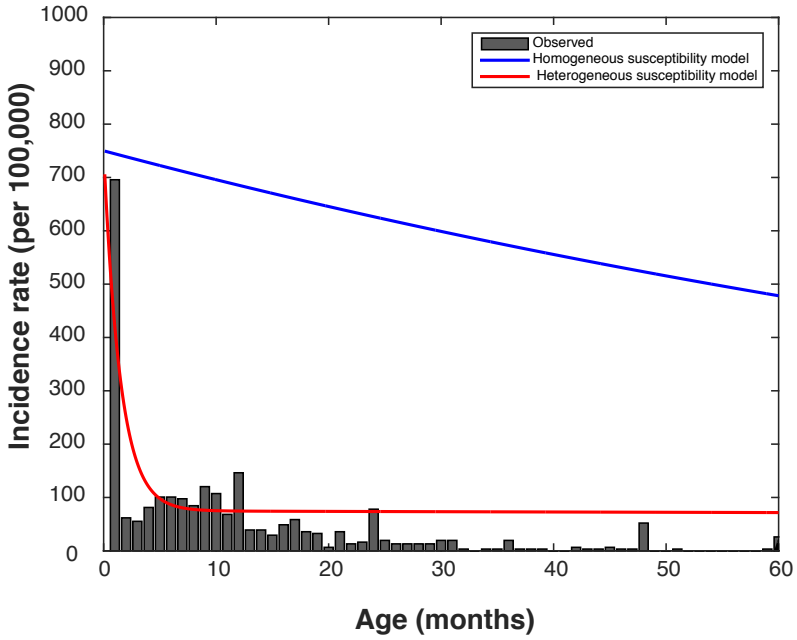

Figure 6.1 Incidence of BSI in children aged < 5 years as predicted by the age-dependent models assuming a population with homogeneous susceptibility to BSI (blue) and a heterogeneous population with two groups with distinct susceptibility levels (red) fitted on a bar plot of observed incidence of BSI..

The model assuming heterogeneous susceptibility was therefore adopted to

estimate incidence over the whole age range. It was assumed that once individuals moved into adulthood (≥ 16 years, they lost the immunity to infection and became susceptible again. A susceptibility factor α was introduced to account for the loss of immunity (primarily due to HIV) in the adults and that individuals in this subpopulation were susceptible to reinfection. However, because the dataset was divided into two age subgroups (< 16 years old) and (≥ 16 years old) with each age group having differential susceptibility, four compartments of susceptibility were created: $S_1$and S_2 for children with susceptibility risk factors $\delta_1$and δ_2; S_3 and S_3 for adults with susceptibility risk factors $\delta_1$and δ_2, respectively. Children moved into adulthood upon reaching age 16 years at rate $\eta = \frac{1}{16}$. Like before, individuals from all compartments of susceptibles got infected at a rate λ and correspondingly moved into compartments I_1, I_2, I_3 and I_4. All infected individuals recovered at the same rate γ. Whilst children acquired immunity after recovering (into R_1 and R_2) from first infection, it was assumed that loss of immunity persisted in adults even after recovery and they became susceptible for reinfection (Figure 6.2).

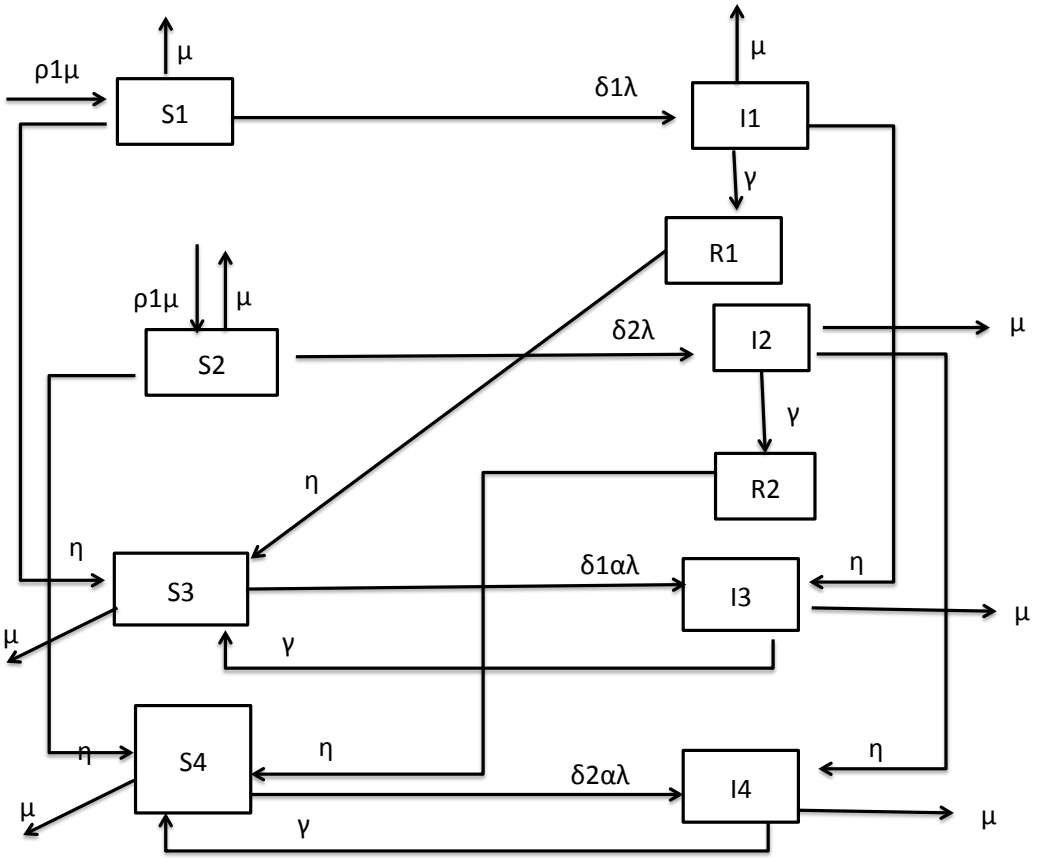

Figure 6.2 A diagrammatic illustration of the heterogeneous model over time

Mathematically, the model was defined by the following set of differential equations:

The model was implemented over time to estimate incidence of BSI over time as described by the following set of differential equations:

$$\frac{dS_1}{dt} = \rho_1\mu - \delta_1\lambda S_1 - \eta S_1 - \mu S_1$$

$$\frac{dS_2}{dt} = \rho_2\mu - \delta_2\lambda S_2 - \eta S_2 - \mu S_2$$

$$\frac{dI_1}{dt} = \delta_1\lambda S_1 - \gamma I_1 - \eta I_1 - \mu I_1$$

$$\frac{dI_2}{dt} = \delta_2\lambda S_2 - \gamma I_2 - \eta I_2 - \mu I_2$$

$$\frac{dR_1}{dt} = \gamma I_1 - \eta R_1 - \mu R_1$$

$$\frac{dR_2}{dt} = \gamma I_2 - \eta R_2 - \mu R_2$$

$$\frac{dS_3}{dt} = \eta(S_1 + R_1) - \alpha\delta_1\lambda S_3 - \mu S_3 - \gamma I_3$$

$$\frac{dS_4}{dt} = \eta(S_2 + R_2) - \alpha\delta_2\lambda S_4 - \mu S_4 - \gamma I_4$$

$$\frac{dI_3}{dt} = \eta I_1 + \alpha\delta_1\lambda S_3 - \gamma I_3 - \mu I_3$$

$$\frac{dI_4}{dt} = \eta I_2 + \alpha\delta_2\lambda S_4 - \gamma I_4 - \mu I_4$$

Thus annual incidence estimated by the model was given by the sum of all new cases from compartments S_1, S_2, S_3 and S_4 as follows:

$Incidence(Y) = \lambda(\delta_1(S_1 + S_3) + \delta_2(S_2 + S_4))$ and hence force of infection:

$$\lambda = \frac{Y}{(\delta_1(S_1 + S_3) + \delta_2(S_2 + S_4)}$$

$$Death\ rate\ (\mu) = \frac{1}{Life\ expectancy}$$

Mean life expectancy for Blantyre city was estimated at 50 years. Hence

$$\mu = 0.02$$

All other parameters are as previously described and are summarised in Table 6.1.

Table 6.1 Summary of model parameters

Parameter	Description	Estimated value
$\boldsymbol{\lambda}$	Force of infection	$\frac{Y}{(\delta_1(S_1+S_3)+\delta_2(S_2+S_4)}$
$\boldsymbol{\mu}$	Death rate	0.017
$\boldsymbol{\delta_1}$	Susceptibility risk factor for individuals in S_1 and S_3	0.98
$\boldsymbol{\delta_2}$	Susceptibility risk factor for individuals in S_2 and S_4	$\frac{1-\rho_1\delta_1}{\rho_2}$
$\boldsymbol{\rho_1}$	Proportion of individuals born with susceptibility risk factor δ_1	0.2
$\boldsymbol{\rho_2}$	Proportion of individuals born with susceptibility risk factor δ_2	0.8
$\boldsymbol{\gamma}$	Recovery rate	52
$\boldsymbol{\eta}$	Rate at which children (≤ 16) years old move into adulthood	$1/16$
$\boldsymbol{\alpha}$	Susceptibility enhancing factor in adults	0.002

The system was implemented in MATLAB.

In general the model showed that incidence of *E. coli* BSI has had a 'relatively' smooth decline in the past two decades and predicts this trend to continue in the years to come if other factors remain unchanged. Although, trend of the observed incidence was characterised by fluctuations, it generally declined with observed values close to the model predicted values, especially after 2006. Whilst observed incidence has been below the model predicted incidence after 2006, it begun to

increase again from2012, going above the model predicted incidence in 2016.

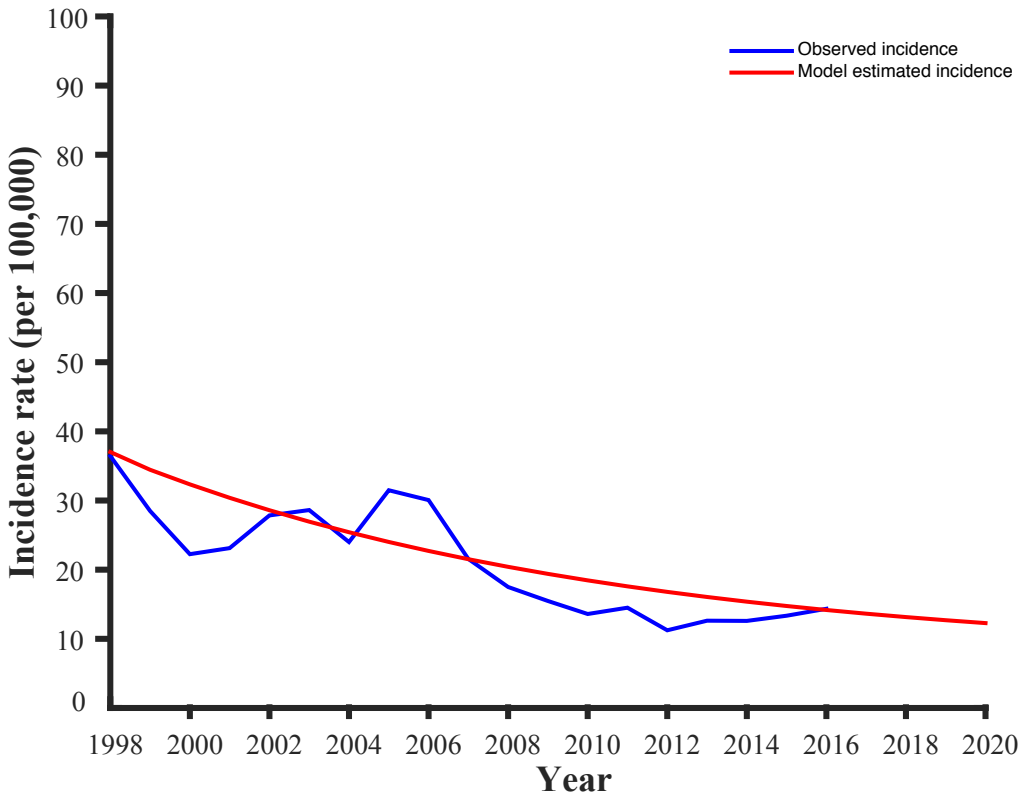

Figure 6.3 Incidence of *E. coli* BSI in children and adults in Blantyre over time estimated by the heterogeneous model (red) fitted against incidence calculated from observed BSI cases (blue).

6.6 Discussion

Infectious disease mathematical models have predominantly been developed with the assumption that the populations under study are homogeneous. However, it is increasingly being appreciated that this is typically not the case and that this is a flawed assumption. A few studies have shown for example that, individuals in a population have heterogenous susceptibility to HIV (Nagelkerke et al., 2009). In this chapter susceptibility to *E. coli* BSI was considered and it has been shown that there is underlying heterogeneity in individuals' susceptibility to BSI.

Taking this heterogeneity into account, a mathematical model was developed to estimate the incidence of *E. coli* BSI in Blantyre. This model does not take into account the contributions of other factors such as malaria, and malnutrition, which as discussed in Chapter Three, drive the spread of BSI or the intervention that coincided with the declining incidence of BSI after 2006. However, their effects are broadly captured by the heterogeneities in susceptibility and by the parameter controlling the decay in incidence. The data presented in other chapters of this thesis has great potential to inform further developments of the model to evaluate the role or effectiveness of various factors such as HIV/AIDS, malnutrition changes in the *E. coli* population structure or interventions such as vaccine introductions in BSI. The model can also be extended to one that would separate DRI from drug-susceptible BSI and make predictions about patterns of DRI in Blantyre.

Between 2000 and 2006 observed incidence increased at a faster rate than the model had estimated where as after 2006 the observed incidence declined more

quickly than the model predicted incidence. This phenomenon might have resulted from the numerous public health interventions that successfully reduced the impact of the factors that had been driving BSI including malaria, HIV and malnutrition (Chapter Three). The observed incidence of *E. coli* BSI has however slowly begun to increase again since 2012, moving above estimated incidence as of 2016. This increase could signal some changes in either in human or *E. coli* populations that are favouring further spread of infection. As noted in Chapter Three, there has been an increase in ESBL and fluoroquinolone resistance amongst *E. coli* isolates in Blantyre over the past decade. As discussed in Chapter four, it is possible that the emergence of CTX-M-15 producing and highly disseminating *E. coli* ST131 clone has been accompanied by its expansion and that it is now causing more BSI than before. It is also possible that there have been changes in the mediators of BSI in the human population that have begun to push the *E. coli* BSI upwards. In order to clearly explain the recent increasing incidence there is a need to develop the model further and evaluate the exact contributions of each of the factors that drive BSI while also modelling separately incidence of BSI caused by ESBL producing and fluoroquinolone resistant *E. coli*.

6.7 Limitations

This chapter has laid the foundation for future model structures to investigate DRI. Comparisons between model outputs and data are crude adjustments. Future developments should adopt established procedures for parameter estimation, such as maximum likelihood or Bayesian inference. Inference for dynamical systems has

been an especially productive area of research over the last decade, and several efficient software packages are now available (for example, Coelho *et al.,* King *et al*., and Toni *et al*. (Toni et al., 2009; Coelho et al., 2011; King, 2016)

6.8 Conclusion

In summary, the overall aim of this work was to estimate the rates at which individuals are acquiring DRI in Blantyre and to evaluate the drivers of its spread. Whilst this is yet to be achieved, the work presented here has established both that there is heterogeneity in susceptibility to BSI in the population in this setting and that this heterogeneity can be accounted for in modelling infection dynamics.

Chapter seven: Concluding remarks and future direction

7.1 Overview

This thesis was carried out to investigate AMR in Blantyre by describing trends in BSI and AMR from longitudinal blood culture dataset, identifying the genetic determinants of AMR in *E. coli* and *K. pneumoniae* isolates and determining the association between the genetic determinants of AMR and the population structures of the two species. To achieve these objectives studies utilising epidemiological and genomic approaches were undertaken. This chapter summarises and discusses the major findings of the studies.

7.2 Summary of Key findings

7.2.1 Decline of bloodstream infection in Blantyre

Analysis of the surveillance blood culture dataset has shown that BSI has substantially declined in Blantyre. This decline has come at a time when there have been improvements in community management of malnutrition following improved food security, increased access to ART and a decline in malaria prevalence and the introduction PCV13.

7.2.2 Emergence and expansion of ESBL and fluoroquinolone resistant Enterobacteriaceae

The positive story of BSI decline in the past two decades is however overshadowed by the increasing threat of untreatable BSI infection in Blantyre. The majority of Gram-negative causes of BSI and a substantial proportion of Gram-positive pathogens were MDR to first line antimicrobial agents (Chapter Three). As a result, ceftriaxone and ciprofloxacin became the antimicrobial agents of choice for treatment of BSI during the surveillance period of 1998-2016. However, Enterobacteriaceae, particularly *E. coli*, and *Klebsiella* spp. expressing ESBL and fluoroquinolone resistance, have emerged and have been expanding in the last decade (Chapter Three). Whilst detection of isolates with ESBL and fluoroquinolone resistance began before cephalosporins and fluoroquinolones were widely used, a substantial rise in the prevalence of ESBL and fluoroquinolone resistance was seen after the introduction of ceftriaxone and ciprofloxacin into extensive use from 2005.

The observed rise in prevalence of ESBL production and fluoroquinolone resistance suggests that the widespread use of ceftriaxone and ciprofloxacin played a role in driving the rapid spread of ESBL and fluoroquinolone resistance in Blantyre, and highlights the need for robust antimicrobial stewardship. In this setting, where cephalosporins and fluoroquinolone are last resort antimicrobial agents, infections caused by pathogens expressing ESBL and fluoroquinolone resistance could become effectively untreatable. The identification of ESBL-producing and fluoroquinolone-

resistant pathogens, and learning how to arrest their spread is therefore of paramount importance.

7.2.3 The changing molecular epidemiology of ESBL resistance

When cephalosporin and fluoroquinolone resistant Enterobacteriaceae were first detected in the early 2000s, a study was carried out to characterise the molecular determinants of ESBL resistance in Blantyre (Gray, 2006). A wide range of ESBLs were detected with SHV-12, SHV-8 and TEM63 being the most common while CTX-M-15 was the least common ESBL and was only identified in only one *K. pneumoniae* isolate. In this thesis it is noted that CTX-M-15 has now replaced SHV and TEM ESBLs as the most common ESBL type in both *E. coli* and *K. pneumoniae* (Chapters Four & Five). One factor linked with the global dominance of CTX-M-15 ESBL is its ability to disseminate more rapidly than other ESBL types (Bonnet, 2004). The replacement of SHV and TEM ESBLs by CTX-M-15 as the most common ESBL type could therefore be due to the greater ease with which $bla_{\text{CTX-M-15}}$ is able to spread across strains compared to the bla_{SHV} and bla_{TEM} genes (Bonnet, 2004). It is also possible that there have been clonal replacements of *E. coli* and *K. pneumoniae* producing TEM and SHV ESBLs by CTX-M-15 producing clones that are intrinsically fitter.

In many settings CTX-M-15 has been associated with CA infection where as SHV and TEM ESBLs have been mostly associated with hospital epidemics. Here, $bla_{\text{CTX-M-15}}$ has been detected in both community invasive and hospital carriage isolates. QECH has now become a reservoir of CTX-M-15 producing strains. The high frequency of

visits to QECH by individuals, either as patients or guardians to patients, result in high rates colonisations of individuals by CTX-M-15 producing strains, furthering their transmission into the community.

7.2.4 Multiple *E. coli* and *K. pneumoniae* lineages carry genotypes encoding ESBL and fluoroquinolone resistance

This thesis has shown that the population structures of both *E. coli* and *K. pneumoniae* in Malawi are diverse, which is consistent with the findings from elsewhere around the globe. Within this huge diversity ST12, ST131 and ST391 were identified as the most common *E. coli* STs whereas ST14 and ST15 for *K. pneumoniae* were the most common *K. pneumoniae* STs. *E. coli* ST131 and ST12 and *K. pneumoniae* ST14 and ST15 are international clones associated with MDR and virulence. In particular *K. pneumoniae* ST14 is emerging as the dominant ST in SSA (Mshana et al., 2013a) while *K. pneumoniae* ST15 has been a cause of hospital epidemics elsewhere across the globe (Chung The et al., 2015; Moradigaravand et al., 2017). The dominance of these STs in the populations of Malawian *E. coli* and *K. pneumoniae* is therefore an indication of international transmission of these STs. As ST14 and ST15 are mostly known for HA infection it would be of interest to investigate their role in HA infection in Malawi.

The diversity observed in the populations of *E. coli* and *K. pneumoniae* extends to strains associated with ESBL or fluoroquinolone resistance in both *E. coli* and *Klebsiella* isolates. Globally, CTX-M-15production and fluoroquinolone resistance has mostly been by a clone of *E. coli* ST131. It has further been noted that QRDR

where mutations associated with fluoroquinolone resistance occur in *E. coli* are subject to recombination events (Petty et al., 2014). The high diversity CTX-M-15 producing and fluoroquinolone resistant STs observed here therefore, does suggest HGT including recombination are contributing to the interlineage exchange of DNA encoding ESBL and fluoroquinolone resistance enabling more strains to acquire ESBL resistance.

Rapid diagnostics and vaccine development are recommended actions for combatting AMR (Infectious Diseases Society of America et al., 2011). Rapid diagnostics will ensure correct identification of pathogens leading to appropriate prescription of antimicrobial agents. In parallel, vaccines against pathogens associated with AMR will limit the ability of these strains to cause disease, leading to a reduced demand for antimicrobial agents and a lower selection pressure for the spread of AMR. The high diversity of ESBL-producing and fluoroquinolone-resistant *E. coli* and *K. pneumoniae* means that diagnostic tools and vaccines targeting such pathogens need to be broad-based to capture the variety of pathogens in the diverse lineages. It will be a long time before such agents, particularly vaccines, are available, suggesting that the rapid institution of antimicrobial stewardship is needed to mitigate further spread of AMR pathogens in SSA.

From the *E. coli* and *K. pneumoniae* genomes studied, no genotypes of carbapenem resistance were identified. This gives hope that carbapenems would be an effective alternative to cephalosporins and fluoroquinolones. However, carbapenem

resistance could spread rapidly ifollowing emergence as the genetic background for carbapenem resistance is already in place (Chapter Four and Chapter Five).

7.2.5 There is underlying heterogeneity in susceptibility to bloodstream infection in Blantyre

Through a mathematical modelling approach, this thesis has shown that there is substantial heterogeneity in susceptibility to BSI in Blantyre, i.e. individuals in the population are not equally susceptible to BSI. Unlike cases where models are developed under the assumption of homogeneous population, the findings in this thesis will encourage incorporation of heterogeneity into future models. Here, a model incorporating heterogeneity in susceptibility to *E. coli* BSI was developed and shown to estimate observed incidence better than a model that assumed a population with homogeneous susceptibility. Further development of the heterogeneous susceptibility model would be useful to predict the incidence of drug resistant infection and to infer the effect of interventions, such as vaccine introduction on the incidence of BSI.

7.3 Future work

This thesis has generated a number of hypotheses regarding genomic epidemiology and transmission dynamics of ESBL and fluoroquinolone resistant strains that need to be tested by further. Firstly, it is hypothesised that improvements in access to ART, community management of malnutrition, enhanced malaria control

programmes and introduction of vaccines have contributed to the declining in incidence of BSI in Blantyre. There is now a need to evaluate the direct impact of each of those factors upon the incidence of BSI.

The findings in this thesis have also shown that the hospital carriage isolates were not phylogenetically distinct from the community invasive isolates and it has been agued here that QECH is serving as a major reservoir of strains with ESBL and fluoroquinolone resistance for the community. Future studies aiming at unravelling the transmission patterns of ESBL producing Enterobacteriaceae between the hospital and the community will be necessary.

E. coli ST131 has emerged as one of the most important *E. coli* clones around the global. The population structure of *E. coli* ST131 in this thesis has revealed a new clade that is unique to the Malawian *E. coli* population. Furthermore, ST131 was not the dominant ESBL producing clone in this setting. These results lead to the hypothesis that genomic epidemiology of both *E. coli* ST131 and CTX-M-15-producing *E. coli* in Malawi is different to that observed in other regions across the globe. This hypothesis should be investigated by sequencing a much larger sample of *E. coli* from this setting. Furthermore, as there is still a gap in understanding ESBL resistance across SSA, it will be necessary to determine if the genomic epidemiology being observed in Malawi reflects that of the wider SSA region.

7.4 Conclusions

This thesis has described a comprehensive and extensive dataset from a sustained period of bacteraemia surveillance. The data in the thesis highlight the value of longitudinal bacteraemia surveillance coupled with detailed molecular epidemiology in all settings, including low-income settings, to understand the global epidemiology of AMR.

Whilst BSI has declined substantially, the emergence of ESBL-producing and fluoroquinolone-resistant Enterobacteriaceae and regular detection of MRSA pose a threat of untreatable infections in this resource-constrained setting in the near future. The predominance of $bla_{\text{CTX-M-15}}$ as the most common ESBL type in both *E. coli* and *K. pneumoniae* poses a threat of further spread of ESBL production and increased treatment failure. A high diversity of lineages were associated with AMR including ESBL and fluoroquinolone resistance in Malawi in both *E. coli K. pneumoniae* isolates, with no dominance of a single clone observed, suggesting the possibility of a distinct genomic backdrop for ESBL production to that observed elsewhere around the globe.

This thesis has shown that ESBL and fluoroquinolone resistance is rapidly spreading in Malawi across multiple *E. coli* and *K. pneumoniae* lineages that are increasingly causing infection. As cephalosporins and fluoroquinolones remain the last resort antimicrobial agents in this setting, there is an urgent need for action to mitigate this spread.

APPENDICES

Appendix 1: Classification of isolates into subgroups

Organism	Subgroup
Alpha- haemolytic streptococcus	Contaminant
Acinetobacter baumanii	Acinetobacter
Acinetobacter junii	Acinetobacter
Acinetobacter lwoffii	Acinetobacter
Acinetobacter sp	Acinetobacter
Acinetobacter species	Acinetobacter
Acinetobacter species	Acinetobacter
Aerococcus species	Anaerobe
Aeromonas hydrophila	Anaerobe
Aeromonas species	Anaerobe
Aeromonas species	Anaerobe
Alcaligene species	Other Gram-negative rods
B-Haemolytic Streptococcus	Other Streptococci/Enterococci
Bacillus species	Contaminant
Bacteroides fragilis	Other Gram-negative rods
Burkholderia cepacia	Pseudomonas
Candida albicans	Yeast
Candida krusei	Yeast
Candida parapsilosis	Yeast
Candida species	Yeast
Citrobacter braakii	Other Enterobacteriaceae
Citrobacter freundii	Other Enterobacteriaceae
Citrobacter species	Other Enterobacteriaceae
Clostridium perfringenes	Contaminant
Clostridium species	Contaminant
Coagulase negative Staph	Contaminant
Coliform species	Other Gram-negative rods
Corynebacterium propinguum	Contaminant
Corynebacterium ureaticum	Contaminant
Cryptococcus neoformans	Yeast

Cryptococcus species	Yeast
Diphtheroids	Contaminant
Escherichia coli	E. coli
Edwardsiella species	Other Enterobacteriaceae
Enterobacter aerogenes	Other Enterobacteriaceae
Enterobacter agglomerans	Other Enterobacteriaceae
Enterobacter amnigenus	Other Enterobacteriaceae
Enterobacter cloacae	Other Enterobacteriaceae
Enterobacter sakazakii	Other Enterobacteriaceae
Enterobacter species	Other Enterobacteriaceae
Enterococci faecalis	Other Streptococci/Enterococci
Enterococcus faecalis	Other Streptococci/Enterococci
Enterococcus faecium	Other Streptococci/Enterococci
Enterococcus species	Other Streptococci/Enterococci
Enterococcus species	Other Streptococci/Enterococci
Escherichia fergusonii	Other Enterobacteriaceae
Escherichia vukneris	Other Enterobacteriaceae
Escherichia vulgalis	Other Enterobacteriaceae
Escherichia vulneris	Other Enterobacteriaceae
Flavobacteria species	Other Enterobacteriaceae
Gram negative rod	Other Gram-negative rods
Gram positive rod	Other Gram positive rod
Group C streptococcus	Other Streptococci/Enterococci
Group D streptococcus	Other Streptococci/Enterococci
Group F streptococcus	Other Streptococci/Enterococci
Group G streptococcus	Other Streptococci/Enterococci
Haemophilus influenzae	Haemophilus
Haemophilus influenzae type b	Haemophilus
Haemophilus influenzae type c	Haemophilus
Haemophilus parainfluenzae	Haemophilus
Hafnei alvei	Other Enterobacteriaceae
Hafnia alvei	Other Enterobacteriaceae
Histophillus somni	Other Gram-negative rods
Klebsiella	Klebsiella
Klebsiella ornithinolytica	Klebsiella
Klebsiella oxytoca	Klebsiella
Klebsiella pneumoniae	Klebsiella
Klebsiella species	Klebsiella
Kluyvera species	Other Enterobacteriaceae
Lactococcus lactis	Contaminant
Leuconostoc species	Contaminant

Listeria species	Other Gram positive rod
Micrococcus species	Contaminant
Moraxella catarrhalis	Pseudomonas
Moraxella species	Pseudomonas
Morganella morganii	Other Enterobacteriaceae
MRSA	S. aureus
Mycobacterium species	Mycobacterium
Mycobacterium tuberculosis	Mycobacterium
Neisseria gonorrhoeae	Neisseria
Neisseria meningitidis	Neisseria
Neisseria meningitidis, group b	Neisseria
Neisseria meningitidis, w135	Neisseria
Neisseria meningitidis; w135	Neisseria
Niesseria species	Neisseria
Gram negative rods	Other Gram-negative rods
Gram positive cocci	Contaminant
Pantoea species	Other Enterobacteriaceae
Pasteurella species	Other Enterobacteriaceae
Proteus mirabilis	Other Enterobacteriaceae
Proteus rettgeri	Other Enterobacteriaceae
Proteus species	Other Enterobacteriaceae
Proteus vulgaris	Other Enterobacteriaceae
Proteusvulgaris	Other Enterobacteriaceae
Pseudomonas aeruginosa	Pseudomonas
Pseudomonas fluorescens	Pseudomonas
Pseudomonas luteola	Pseudomonas
Pseudomonas oryzihabitans	Pseudomonas
Pseudomonas putida	Pseudomonas
Pseudomonas species	Pseudomonas
Pseudomonas stutzeri	Pseudomonas
Raoultella ornithinolytica	Other Enterobacteriaceae
Raoultella species	Other Enterobacteriaceae
Rhizobium radiobacter	Contaminant
Salmonella typhi	S. Typhi
Salmonella enteritidis	S. Enteritidis
Salmonella species	Salmonella
Salmonella typhimurium	S. Typhimurium
Serratia liquefaciens	Other Enterobacteriaceae
Serratia marscesens	Other Enterobacteriaceae
Serratia oderfera	Other Enterobacteriaceae
Serratia species	Other Enterobacteriaceae

Shigella boydii	Other Enterobacteriaceae
Shigella dysenteriae	Other Enterobacteriaceae
Shigella flexineri	Other Enterobacteriaceae
Shigella sonnei	Other Enterobacteriaceae
Shigella species	Other Enterobacteriaceae
Skin flora	Contaminant
Staphylococcus aureus	S. aureus
Stenotrophomonas maltophilia	Pseudomonas
Streptcococcus species	Other Streptococci/Enterococci
Streptococcus agalactiae	Other Streptococci/Enterococci
Streptococcus pneumoniae	S. pneumoniae
Streptococcus pyogenes	Other Streptococci/Enterococci
Streptococcus species	Other Streptococci/Enterococci
Vibrio cholerae	Vibrio
Vibrio parahaemolyticus	Vibrio
Xanthomonas maltophilia	Pseudomonas
Xanthomonas species	Pseudomonas
Yeast species	Yeast
Yersinia pseudotuberculosis	Other Enterobacteriaceae
Yersinia species	Other Enterobacteriaceae

Appendix 2A Number of blood cultures by subgroup, 1998-2007

	Year									
BC subgroup	**1998**	**1999**	**2000**	**2001**	**2002**	**2003**	**2004**	**2005**	**2006**	**2007**
S. Typhimurium	451	375	421	446	1045	1187	710	737	601	563
S. pneumoniae	356	345	279	159	261	356	289	570	420	329
E. coli	192	151	117	132	159	166	143	193	190	139
S. Typhi	34	14	13	6	6	9	14	20	18	11
S. aureus	102	115	160	128	109	124	106	141	76	49
S. Enteritidis	80	209	247	318	226	106	79	78	53	49
Other Strept /Enterococci	118	107	155	87	106	77	86	160	86	60
Klebsiella	149	121	104	75	83	61	54	50	53	59
Other Enterobacteriaceae	47	29	97	138	88	124	71	64	80	33
Yeast	7	14	10	23	62	64	53	76	120	77
Other Salmonellae	21	39	59	19	115	30	87	32	17	17
Acinetobacter	56	61	38	45	35	25	15	51	51	28
Haemophilus	16	41	48	48	69	18	25	33	30	11
Pseudomonas	33	23	25	17	28	18	29	27	10	11
Other Gram-negtaive rods	4	3	3	8	2	0	20	23	19	5
Neisseria	40	27	6	5	4	4	2	17	15	14
Mycobacterium	0	0	60	0	0	0	0	0	0	0
Other Gram-neg. cocci	10	1	3	1	5	1	3	2	2	0
Anaerobes	5	1	0	1	5	2	1	0	1	1
Vibrio	3	2	0	0	0	0	0	2	2	0
Mixed growth	19	1	1	0	1	0	0	0	0	0
Unknown	19	21	23	17	23	28	28	25	7	4
Contaminants	1348	1429	1893	1203	1701	2907	2536	2430	1508	1398
No growth	5435	4694	4290	4777	4845	5867	6302	8202	6764	6309
Total BC	**8545**	**7823**	**8052**	**7653**	**8978**	**11174**	**10653**	**12933**	**10123**	**9167**

Appendix 2B- Number of blood cultures by subgroup, 2008-2016

	Year								
BC subgroup	**2008**	**2009**	**2010**	**2011**	**2012**	**2013**	**2014**	**2015**	**2016**
S. Typhimurium	443	287	209	185	131	129	119	122	132
S. pneumoniae	166	157	123	177	107	41	31	50	42
E. coli	116	107	98	109	88	103	107	118	132
S. Typhi	16	25	21	69	191	887	780	565	298
S. aureus	72	61	72	73	104	95	94	118	126
S. Enteritidis	42	36	34	22	19	12	13	17	18
Other Strept/Enterococci	43	38	41	81	59	75	71	80	79
Klebsiella spp.	56	43	47	48	40	55	51	48	84
Other Enterobacteriaceae	28	24	36	19	30	61	65	33	65
Yeast	62	45	50	55	43	60	62	55	65
Salmonellae spp.	13	20	21	9	4	7	6	3	4
Acinetobacter	16	20	18	10	12	24	18	4	16
Haemophilus spp.	12	9	13	13	13	12	3	8	12
Pseudomonas spp.	14	6	11	29	57	29	25	28	22
Other Gram-negative rods	12	5	7	14	18	16	7	3	6
Neisseria spp.	8	5	3	2	3	16	11	24	4
Mycobacterium spp	0	0	0	0	0	0	0	0	0
Other Gram-neg. cocci	5	4	0	0	0	0	1	0	0
Anaerobes	1	4	1	1	3	1	2	0	0
Vibrio spp.	0	0	2	0	1	0	0	0	0
Mixed growth	0	0	0	0	0	0	0	0	0
Unknown	1	0	0	0	0	0	0	0	0
Contaminants	1524	1499	1711	1992	2041	2433	2282	2387	2542
No growth	5978	5596	5989	6982	7469	8759	9915	10164	10037
Total BC	**8628**	**7991**	**8507**	**9890**	**10433**	**12815**	**13663**	**13827**	**13684**

Appendix 3 List of *E. coli* isolates sequenced, closest species genome match, and assembly statistics

Lane ID	Sample ID	Total Raw Reads	Top match: Match%	Total Length	Avg Contig Length	N50	Contigs in N50
15966_8#37	A7651	5030534	*E. coli: 49.4*	10508894	7549.49	44744	42
16228_1#49	B9	79642	*E. coli: 52.0*	426	426	426	1
15966_8#50	A40286	4334702	*E. coli: 58.3*	5587432	19400.81	93942	16
15966_8#80	B28	4256968	*E. coli: 58.9*	5404547	61415.31	192412	8
15966_8#76	D32322	4759878	*E. coli: 59.4*	5588592	46962.96	189956	9
16228_1#44	D46760	4520844	*E. coli: 59.7*	5783150	30120.57	108336	15
15966_8#6	BKQA8N	4832992	*E. coli: 59.9*	5609738	43151.83	182674	11
15966_8#71	C4	4891580	*E. coli: 60.1*	5291658	41997.29	181330	9
15966_8#32	D3420	4727328	*E. coli: 60.4*	5457019	34980.89	188255	10
15966_8#78	D3871	4651836	*E. coli: 60.9*	5431645	34161.29	149291	13
15966_8#2	D3787	3907992	*E. coli: 61.1*	5387283	33254.83	150578	13
15966_8#25	A7503	4335038	*E. coli: 61.7*	5326194	26899.97	156280	11
15966_8#44	A25576	4368046	*E. coli: 61.9*	5428549	31745.90	129069	14
15966_8#82	D49086	4009616	*E. coli: 62.1*	5431220	42102.48	160139	11
15966_8#81	C20	4231500	*E. coli: 62.2*	5339050	36320.07	135868	11
15966_8#59	A32883	4742876	*E. coli: 62.3*	5301863	34880.68	164400	11
15966_8#12	1010805	4338346	*E. coli: 62.4*	5204957	29573.62	87988	19
15966_8#69	D40059	4552514	*E. coli: 62.5*	5341539	58060.21	155469	8
15966_8#4	C1289	4011682	*E. coli: 62.7*	5276857	36392.12	221661	9
15966_8#70	D43713	4645450	*E. coli: 62.9*	5341172	54501.76	162589	11
15966_8#29	A5175	4685120	*E. coli: 63.1*	5294462	40109.56	180364	12
15966_8#64	C29	5529764	*E. coli: 63.1*	5271826	58575.84	191186	9
15966_8#8	D29253	4400748	*E. coli: 63.1*	5440270	56085.26	203704	9
15966_8#42	D36115	4398010	*E. coli: 63.3*	5213958	42737.36	265285	7
15966_8#35	10151	4391630	*E. coli: 63.4*	5350547	52975.71	189034	9
15966_8#47	D25641	4818522	*E. coli: 63.6*	5346456	56877.19	191459	10
16228_1#42	10140	5210926	*E. coli: 63.6*	5348508	53485.08	180964	9
15966_8#43	D2209	4279218	*E. coli: 64.0*	5289536	51858.20	184080	10
15966_8#84	D48799	4270816	*E. coli: 64.1*	5232386	63040.80	204504	9
15966_8#16	D40034	4496170	*E. coli: 64.3*	5336252	52316.20	181853	8
15966_8#56	4600	5231072	*E. coli: 64.3*	5394345	50414.44	215073	7
16228_1#43	D42544	4495808	*E. coli: 64.4*	5556477	59111.46	188605	9
15966_8#22	BKR406	4724706	*E. coli: 64.5*	5164557	30741.41	84814	20
15966_8#13	BHA15G	4478474	*E. coli: 64.6*	5290632	66132.90	313280	7

15966_8#62	2558	4373852	*E. coli: 64.6*	5213209	32180.30	141239	9
15966_8#1	A39011	4370028	*E. coli: 64.8*	5273808	53270.79	205901	7
16228_1#46	A45016	4449422	*E. coli: 65.2*	4972999	63756.40	302059	6
15966_8#27	A38084	4570168	*E. coli: 65.6*	5108748	49122.58	139415	8
15966_8#24	B12381	4638042	*E. coli: 65.8*	5112808	40902.46	127923	7
15966_8#58	C12359	4267864	*E. coli: 65.8*	5513173	26762.98	152004	12
15966_8#68	C15	4411414	*E. coli: 65.9*	5331258	41327.58	219011	9
15966_8#10	522A	4091108	*E. coli: 66.0*	5461868	63510.09	205687	7
15966_8#17	BKQ5JN	4287670	*E. coli: 66.0*	5380861	65620.26	177456	6
15966_8#30	D4531	4762958	*E. coli: 66.0*	4953068	43447.96	121138	13
15966_8#33	C10382	4321038	*E. coli: 66.0*	5319183	43245.39	131769	14
15966_8#55	3524	4838170	*E. coli: 66.1*	5170502	47435.80	190806	8
15966_8#77	C14	4496520	*E. coli: 66.1*	5329072	50753.07	166888	9
15966_8#54	A36140	4563326	*E. coli: 66.4*	5101616	51016.16	247618	7
15966_8#63	A27	4680432	*E. coli: 66.4*	5334528	48495.71	175780	10
15966_8#73	A35440	4256024	*E. coli: 67.2*	5153397	23531.49	65279	26
15966_8#39	A36329	4786292	*E. coli: 67.7*	5362879	50593.20	165106	9
15966_8#51	4464	4466068	*E. coli: 67.7*	5179118	35718.06	109887	17
15966_8#38	10129	5083566	*E. coli: 67.8*	4951467	72815.69	261762	6
15966_8#14	BKQ7M8	4627100	*E. coli: 68.6*	5312488	43190.96	160639	10
15966_8#11	A45214	3669562	*E. coli: 68.9*	5284968	55051.75	175071	9
15966_8#79	1014142	4595946	*E. coli: 69.1*	5062819	42190.16	134448	11
16228_1#45	A48349	4585460	*E. coli: 69.8*	5361198	55270.08	175891	10
15966_8#19	3361	4308016	*E. coli: 70.1*	5503294	39592.04	280302	8
15966_8#23	D3475	4721348	*E. coli: 70.1*	5283543	48472.87	126471	14
15966_8#28	D3275	4474174	*E. coli: 70.6*	5270756	58563.96	180266	10
15966_8#41	9597	4724184	*E. coli: 71.3*	5094548	92628.15	206791	9
15966_8#74	D3454	3861496	*E. coli: 71.6*	7257834	1239.60	2234	883
15966_8#72	A16	5264024	*E. coli: 71.9*	5250580	39777.12	113555	13
15966_8#75	C33b	4394740	*E. coli: 71.9*	5303054	43467.66	122520	11
15966_8#18	A7898	4286372	*E. coli: 72.0*	5480078	30276.67	127272	15
15966_8#49	A38988	4800982	*E. coli: 72.5*	5382272	25876.31	108438	17
15966_8#3	B1PG3U	4195422	*E. coli: 72.8*	5172184	32735.34	87822	17
15966_8#20	1016997	4440000	*E. coli: 73.0*	5140984	9963.15	123743	10
15966_8#66	D39719	4175868	*E. coli: 73.0*	5163387	55520.29	226049	9
15966_8#53	C14036	4528054	*E. coli: 73.5*	5244671	27174.46	87604	20
15966_8#48	D29454	4681176	*E. coli: 74.0*	5113300	44463.48	183101	9
15966_8#61	9693	4443872	*E. coli: 74.8*	5224453	32858.19	114505	14
15966_8#46	D25640	4654940	*E. coli: 75.1*	5282620	31823.01	143794	11
15966_8#9	1016948	4283526	*E. coli: 75.7*	5155016	59253.06	244350	7
15966_8#52	B9222	4173864	*E. coli: 76.2*	5063584	58202.11	240689	7

15966_8#34	2228	3860204	*E. coli: 77.3*	5070388	32712.18	108237	16
15966_8#45	D37334	4758584	*E. coli: 78.3*	5021537	25107.69	78986	21
15966_8#65	A333	4718882	*E. coli: 78.3*	4981825	49818.25	154051	9
15966_8#15	BKR1Z7	4656302	*E. coli: 79.1*	4903006	86017.65	235975	7
15966_8#67	C30	4762590	*E. coli: 79.6*	5185692	63240.15	201161	9
15966_8#26	C301	4241544	*E. coli: 79.7*	5015203	52791.61	147873	10
15966_8#36	8728	4226264	*E. coli: 79.9*	4897747	33318.01	85461	19
15966_8#7	BHAIAI	4249344	*E. coli: 80.9*	4854690	52768.37	122489	15
16228_1#48	A3b	5414942	*E. coli: 81.2*	4937459	58087.75	143228	11
15966_8#85	B3	4208682	*E. coli: 82.6*	4959972	49599.72	150867	11
15966_8#5	1012184	4456998	*E. coli: 82.7*	4938419	51983.36	118104	13
15966_8#40	B9070	4715208	*E. coli: 82.8*	4885674	46977.63	150391	12
15966_8#31	2473	4530228	*E. coli: 85.8*	4919531	58565.85	210321	8
15966_8#57	4604	4600260	*K. pneumoniae: 55.4*	5879774	77365.45	294799	7
16228_1#47	A4	5173042	*K. pneumoniae: 56.2*	5739948	78629.42	253789	7
15966_8#21	BKQ79K_2	4270538	*K. pneumoniae: 60.4*	5379195	103446.06	350097	6
15966_8#83	D39846	4672184	*P. mirabilis: 44.1*	9673722	2089.81	72274	40
15966_8#60	B1014	4039968	*C. rodentium: 6.5*	4834178	142181.71	490386	4

Appendix 4 Sequenced *E. coli* isolates metadata

Strain_ID	Source	Age group	ST	Phylo group	AMP	CHL	COTRIM	GENT	CEFT	CIPR
A39011	Blood	Adult	73	B2	R	R	S	S	S	S
522_A	Blood	Child	362	F	R	R	S	S	S	S
A45214	Blood	Adult	69	D	R	R	S	R	S	S
1010805	Blood		504	B2	S	S	R	S	S	S
BHA15G	Blood	Adult	12	B2	R	S	R	S	S	S
BKQ7M8	Blood	Child	69	D	R	R	R	S	S	S
BKR1Z7	CSF	Child	196	B1	S	S	S	S	S	S
D40034	Blood	Child	391	D	R	R	S	R	R	R
BKQ5JN	Blood	Child	391	D	R	R	R	R	R	R
A7898	Blood	Adult	223	B1	S	S	S	S	S	S
3361	CSF		335	F	S	R	S	S	S	S
D3787	Blood	Child	12	B2	R	R	S	R	R	S
BKR406	CSF	Child	59	F	R	R	R	S	S	S
D3475	Blood	Child	69	D	R	S	R	S	S	S
B12381	CSF	Adult	5148	D	R	S	S	S	S	S
A7503	Blood	Adult		B2	R	R	R	S	S	S
C301	CSF	Child	652	A	R	R	R	S	S	S
A38084	Blood	Adult	5148	D	R	S	R	S	S	R
D3275	Blood	Child	69	D	R	R	R	S	s	S
A5175	Blood	Adult	12	B2	R	R	R	S	S	S
B1PG3	Blood	Child	52	A	S	S	R	S	S	S
D4531	Blood	Child	15	B2	R	R	S	S	S	S
2473	Blood	Child	167	A	R	S	R	S	S	S
D3420	Blood	Child	131	B2	R	R	R	R	S	S
C10382	CSF	Child	1567	D						
2228	CSF	Child	10	A	R	R	S	S	S	S
10151	CSF	Child	131	B2	R	R	S	R	S	S
8728	CSF	Child	4656	A	R	S	S	S	S	S
10129	CSF	Child	636	B2	S	S	S	S	S	S
A36329	Blood	Adult	69	D	R	R	R	S	S	S
C1289	CSF	Child	131	B2	R	R	S	R	S	S
B9070	CSF	Adult	10	A	R	R	R	S	S	R
9597	CSF	Child	1567	D	R	R	R	S	S	S
D36115	Blood	Child	12	B2	R	R	R	S	S	S
2209	Blood	Child	12	B2	R	R	R	S	S	S
A25576	Blood	Child	405	F						

D37334	Blood	Child	48	A	R	R	R	S	S	S
D25640	Blood	Child	10	A	R	R	S	S	S	S
D25641	Blood	Child	131	B2	R	R	R		S	S
D29454	Blood	Child	1084	B1						
A38988	Blood	Adult	10	A	R	R	R	S	S	S
1012184	Blood	Adult	361	A	R	R	S	R	R	R
A40286	Blood	Adult	405	F	R	R	R	S	S	S
4464	CSF	Child	62	F	S	S	S	S	S	S
B9222	CSF	Adult	500	B1	R	R	R	S	S	S
C14036	CSF	Child	1312	A	R	S	S	S	S	S
A36140	Blood	Adult	73	B2	S	S	S	S	S	S
3524	Blood		12	B2	R	R	S	S	S	S
4600	Blood		73	B2	R	R	S	S	S	S
C12359	CSF	Child	38	D	R	R	S	S	S	S
A32883	Blood	Adult	131	B2	R	R	R	R	S	S
BKQA8N	Blood	Child	131	B2	R	R	R	R	R	R
9693	Blood	Child	10	A	R	R	S	S	S	S
2558	Blood	Child	12	B2	R	R	S	S	S	S
A27	RS	Adult	648	F	R	R	R	R	R	R
C29	CSF	Adult	131	B2						
A333	Blood	Adult	448	B1	R	S	R	R	R	R
D39719	Blood	Child	410	A	R	R	S	R	S	R
C30	RS	Adult	167	A	R	R	R	R	R	R
C15	CSF	Adult	648	F	R	R	R	R	R	R
D40059A	Blood	Child	391	D	R	R	S	R	R	R
BHAIAI	Blood	Child		A						
D43713	Blood	Child	2279	B2	R	R	S	R	S	S
C4	CSF	Adult	131	B2	R	R	R	R	R	R
A16	RS	Adult	448	B1	R	S	R	R	R	R
A35440	Blood	Adult	280	D	S	S	S	S	S	S
C33B	CSF	Adult	167	A	R	R	R	R	R	R
D32322	Blood	Child	131	B2	R	R	S	R	S	S
C14	CSF	Adult	391	D						
D3871	Blood	Child	12	B2	R	R	R	R	R	S
1014142	Blood		1163	F	R	R	S	R	R	S
D29253	Blood	Child	131	B2	R	S	S	S	S	S
B28	RS	Adult	131	B2	R	R	R	R	R	R
C20b	RS	Adult	648	F	R	S	R	R	R	R
D49086	Blood	Child	131	B2	R	R	S	R	S	S
D48799	Blood	Child	131	B2	R	R	S	R	S	S
B3	RS	Adult	617	A	R	S	R	R	R	R

1016948	CSF	Child	977	B1	S	R	S	R	R	R
BKQ79K	Blood	Child	4358	A	R	S	S	S	S	S
10276	Blood	Child	501	F	R	R	S	S	S	
D26076	Blood	Child		D	R	R	R	S	S	S
A7881	Blood	Adult	12	B2	R	R	R	S		
D45621	Blood	Child	354	F	R	R	R	R	R	S
D4275	Blood	Child		D	R	R	R	S	S	S
A44893	Blood	Adult	1286	A	R	R	R	S	S	S
A1a	RS	Adult	399	A	R	R	R	R	R	R
10140	CSF		131	B2	R	R	S	R	S	S
D42544	Blood	Child		D	R	R	S	R	R	R
D46760	Blood	Child	2141	F	R	R	S	R	R	S
A48349	Blood	Adult	38	D	R	R	S	R	S	R
A45016	Blood	Adult	636	B2	R	S	R	S	S	S
A3b	RS	Adult	617	A	R	R	S	R	R	R
D33237	Blood	Child	120	A	S	S	S	S	S	S

Appendix 5 List of *K. pneumoniae* isolates sequenced, closest species genome matches and assembly statistics

Lane	Sample_ID	Top match: Match%	Total Length	No. Contigs	Avg Contig Length	N50	Contigs in N50
15966_8#87	522B	*K. pneumoniae: 53.7*	5766474	55	104844	292753	5
15966_8#88	B1874	*K. pneumoniae: 92.1*	5490924	37	148403	347270	6
15966_8#89	1011136	*K. pneumoniae: 59.3*	5317672	51	104268	216739	7
15966_8#90	2669	*K. pneumoniae: 45.0*	9206813	4539	2028	27922	77
15966_8#91	BKR469	*K. pneumoniae: 57.1*	5691981	62	91806	225484	6
15966_8#92	BKQ653	*A. baumannii: 26.9*	3774868	50	75497	179634	6
15966_8#93	BKQ3KY	*K. pneumoniae: 63.8*	5126380	28	183085	369848	4
15966_8#94	BKQ79K_1	*E. coli: 77.0*	5020018	210	23904	78702	19
15966_8#95	BKQ2QU	*K. pneumoniae: 55.9*	5528251	59	93699	598889	3
16228_1#1	BKQAH1	*K. pneumoniae: 10.3*	5749361	28	205334	457830	4
16228_1#10	D25672	*C. rodentium: 5.9*	4808626	23	209070	537959	4
16228_1#11	13	*K. pneumoniae: 53.9*	5404431	69	78325	434141	5
16228_1#12	C773	*S. enterica: 90.3*	4754130	17	279654	477400	3
16228_1#13	D25597	*K. pneumoniae: 55.2*	5883009	77	76402	315621	7
16228_1#14	3712 1	*K. pneumoniae: 56.2*	5707917	53	107696	379353	6
16228_1#15	8407	*K. pneumoniae: 59.4*	5374896	47	114359	358949	6
16228_1#16	10276	*E. coli: 68.8*	5281674	105	50301	211874	9
16228_1#17	D26076	*E. coli: 67.4*	5477833	101	54235	170807	11
16228_1#18	776	*K. pneumoniae: 55.8*	5719931	70	81713	384848	3
16228_1#19	D32265	*K. pneumoniae: 61.1*	5379430	30	179314	342703	5
16228_1#2	D36068	*K. pneumoniae: 53.6*	5717347	69	82860	360830	5
16228_1#20	A7881	*E. coli: 61.1*	5367711	123	43640	181395	10
16228_1#21	A44754	*K. pneumoniae: 59.7*	5491510	69	79587	357732	6
16228_1#22	D45621	*E. coli: 65.1*	5206131	62	83969	326072	7
16228_1#23	D50395	*K. pneumoniae: 58.1*	5620716	82	68545	335661	6
16228_1#24	D4275	*E. coli: 73.9*	5030609	50	100612	260601	6
16228_1#25	D49363	*B. thuringiensis: 50.7*	10944428	125	87555	315738	11
16228_1#26	D47091	*K. pneumoniae: 59.0*	5738423	110	52167	237328	7
16228_1#27	D1205	*K. pneumoniae: 20.3*	13794371	783	17617	66736	51
16228_1#28	C28	*K. pneumoniae: 55.6*	5546990	55	100854	370889	5
16228_1#29	C21	*K. pneumoniae: 57.6*	5554961	121	45909	139068	13
16228_1#3	D41222	*K. pneumoniae: 55.4*	5686157	80	71077	294511	7
16228_1#30	A28	*K. pneumoniae: 60.0*	5536257	36	153785	429016	5
16228_1#31	C18	*K. pneumoniae: 53.5*	5671080	127	44654	228115	8

16228_1#32	C24b	*K. pneumoniae: 52.2*	5663163	66	85806	338467	7
16228_1#33	A44893	*E. coli: 71.1*	5219699	191	27328	106000	17
16228_1#34	1007011	*K. pneumoniae: 58.4*	5669749	75	75596	247449	7
16228_1#35	D27401	*K. pneumoniae: 9.6*	6239807	72	86664	400135	6
16228_1#36	1023547	*K. pneumoniae: 56.3*	5813138	95	61191	330737	6
16228_1#37	1026453	*K. pneumoniae: 14.2*	6578346	4316	1524	205518	11
16228_1#38	BKQ2KD	*K. pneumoniae: 57.9*	5539695	75	73863	204883	8
16228_1#39	D50395	*K. pneumoniae: 86.7*	5616469	84	66863	247706	7
16228_1#4	BHD3XR	*K. pneumoniae: 91.1*	5455292	26	209819	708913	4
16228_1#40	D25466	*K. pneumoniae: 57.4*	5548538	56	99081	318246	6
16228_1#41	A1a	*E. coli: 74.6*	5201081	167	31144	84058	19
16228_1#5	BKR2U6	*K. pneumoniae: 54.9*	5588368	65	85975	209424	9
16228_1#50	C32	*K. pneumoniae: 58.9*	5512052	86	64094	368203	6
16228_1#51	3208	*K. pneumoniae: 57.4*	5745021	67	85747	289266	7
16228_1#52	D5105	*K. pneumoniae: 47.5*	5956644	166	35883	149809	13
16228_1#53	D37100	*K. pneumoniae: 59.5*	5372427	86	62470	220575	10
16228_1#54	D4743	*K. pneumoniae: 55.4*	5359651	32	167489	408125	5
16228_1#55	D25884	*K. pneumoniae: 53.5*	5822050	120	48517	133451	14
16228_1#56	8193	*K. pneumoniae: 56.6*	5890457	57	103341	296288	8
16228_1#57	A45755	*K. pneumoniae: 56.9*	5393714	47	114759	380971	6
16228_1#58	D4888	*K. pneumoniae: 59.5*	5484926	77	71232	349059	6
16228_1#59	D35263	*K. pneumoniae: 56.3*	5719197	75	76256	366159	5
16228_1#6	A40483	*K. pneumoniae: 51.8*	5663446	158	35845	370900	6
16228_1#60	B10140	*K. pneumoniae: 62.3*	5230653	238	21977	35793	42
16228_1#61	C1408	*K. pneumoniae: 58.3*	5499864	47	117018	363091	6
16228_1#62	D33237	*E. coli: 72.3*	4713852	282	16716	47506	35
16228_1#63	D44912	*K. pneumoniae: 60.5*	5532322	39	141854	352062	6
16228_1#64	A28005	*K. pneumoniae: 59.7*	5701202	61	93462	350528	6
16228_1#65	D30014	*K. pneumoniae: 58.3*	5577609	70	79680	374774	6
16228_1#66	D46208	*K. pneumoniae: 60.3*	5420352	58	93454	370178	5
16228_1#67	D40443	*K. pneumoniae: 56.7*	5816385	91	63916	362938	6
16228_1#68	D42680	*K. pneumoniae: 55.4*	5818448	162	35916	150900	11
16228_1#69	D40180	*K. pneumoniae: 60.5*	5553852	66	84149	245404	6
16228_1#7	6	*K. pneumoniae: 59.5*	5469129	64	85455	410144	5
16228_1#70	D3538	*K. pneumoniae: 53.7*	5974058	116	515001	199660	8
16228_1#71	D29665	*K. pneumoniae: 56.9*	5869136	74	79313	253093	9
16228_1#72	D47244	*K. pneumoniae: 59.4*	5463963	44	124181	407611	5
16228_1#73	D51871	*K. pneumoniae: 55.5*	5561367	112	49655	368469	5
16228_1#74	C24a	*K. pneumoniae: 52.2*	5589840	55	101633	337278	7
16228_1#75	C23	*K. pneumoniae: 53.8*	5671159	83	68327	608990	3
16228_1#76	C35	*K. pneumoniae: 60.0*	5474345	118	46393	114482	16

16228_1#77	A43903	*K. pneumoniae: 57.6*	5687055	61	93230	320117	7
16228_1#78	D51871	*K. pneumoniae: 83.5*	5542232	146	37960	463273	5
16228_1#79	1022430	*K. pneumoniae: 56.0*	5815543	106	54864	372167	6
16228_1#8	BKQAV6	*K. pneumoniae: 57.3*	5538236	57	97162	272986	6
16228_1#80	D39172	*K. pneumoniae: 56.7*	5844814	123	47519	188592	9
16228_1#81	D49363	*K. pneumoniae: 58.5*	5549172	51	108807	350489	6
16228_1#82	D30074	*K. pneumoniae: 57.5*	5435740	67	81130	269920	7
16228_1#83	10038396	*K. pneumoniae: 56.9*	5597331	76	73649	226058	10
16228_1#84	D53369	*K. pneumoniae: 57.7*	5759202	140	41137	174183	10
16228_1#85	1027734	*K. pneumoniae: 58.8*	5538550	75	73847	275484	6
16228_1#86	D572	*B. thuringiensis: 62.9*	6004333	60	100072	434543	5
16228_1#9	A44810	*K. pneumoniae: 63.8*	5206046	26	200233	420628	3

Appendix 6 Metadata for sequenced *K. pneumoniae* isolates

Isolate	Year	Source	Age group	SC	ST	AMP	CHL	COT	CTX	CIP	GENT
D46208	2008	Blood	Child	KpI	ST4	R	R	R	R	R	R
4604	1997	CSF		KpI	ST14	R	S	S			
D25597	2004	Blood	Child	KpI	ST14	R	R	R	S	S	R
1023547	2011	Blood	Adult	KpI	ST14	R	R	R	R	S	R
8193	1998	Blood	Adult	KpI	ST14						
D44912	2008	Blood	Child	KpI	ST14	R	R	R	R	S	R
D3538	1999	Blood	Child	KpI	ST14	R	R	R			R
D29665	2005	Blood	Child	KpI	ST14	R	R	R	S	S	R
1022430	2011	Blood		KpI	ST14	R	R	R	R	R	R
D39172	2006	Blood	Child	KpI	ST14	R	R	R	R	S	R
D53369	2009	Blood	Child	KpI	ST14	R	R	R	R	S	R
1007011	2010	Blood	Adult	KpI	ST15	R	R	R	R	R	S
D25466	2004	Blood	Child	KpI	ST15	R	R	R	R	R	R
C24a	2009	RS	Adult	KpI	ST15	R	R	R	R	R	R
522B	1996	Blood		KpI	ST17	R	R	R	S	S	S
D30074	2005	Blood	Child	KpI	ST20	R	S	S	S	S	S
C21	2009	RS	Adult	KpI	ST25	R	R	R	R	R	R
D4888	2000	Blood	Child	KpI	ST25	R	R	R	S	S	S
D40180	2006	Blood	Child	KpI	ST34	R	R	R	S	S	R
BKQ2ZY	2012	Blood	Child	KpI	ST37	R	S	R	S	S	S
C18	2009	RS	Adult	KpI	ST39	R	R	R	R	S	R
D32265	2005	Blood	Child	KpI	ST45	R	R	R	S	S	S
3712	1997	Blood		KpI	ST48	R	R	R	S	S	R
3208	1997	Blood	Adult	KpI	ST48	R	R	R	S	S	R
D49363	2008	Blood	Child	KpI	ST48	R	R	R	R	R	S
D5105	2000	Blood	Child	KpI	ST73	R	R	R	S	S	R
D51871	2009	Blood	Child	KpI	ST101	R	R	R	R	S	R
D51871B	2009	Blood	Child	KpI	ST101						
BKQ3KY	2012	Blood	Child	KpI	ST106						
C24b	2009	RS	Adult	KpI	ST107	R	R	R	R	S	R
D50395	2009	Blood	Child	KpI	ST198	R	R	R	R	S	R
D50395	2009	Blood	Child	KpI	ST198	R	R	R	R	S	R
C1408	1999	CSF	Child	KpI	ST218	R	R	R	S	S	R
A45755	2007	Blood	Adult	KpI	ST231	R	R	R	R	R	S
D28005	2004	Blood	Child	KpI	ST268	R	R	R	S	S	R

A43903	2006	Blood	Adult	KpI	ST268						
D47244	2008	Blood	Child	KpI	ST276	R	R	R	S	S	R
D4743	2000	Blood	Child	KpII	ST283	R	S	S	S	S	S
D25884	2004	Blood	Child	KpI	ST290						
D37100	2006	Blood	Child	KpI	ST297	R	R	R	S	S	S
BKQ2KD	2012	Blood	Child	KpI	ST307	R	R	R	R	R	R
A4		RS	Adult	KpI	ST307	R	R	R	R	R	R
BKR2U6	2013	CSF	Child	KpI	ST340	R	R	R	R	R	R
BKR469	2014	Blood	Child	KpI	ST348	R	R	R	S	R	R
BKQAV6	2014	Blood	Child	KpI	ST372						
1011136	2011	Blood		KpI	ST383	R	S	S	S	S	S
D41222	2007	Blood	Child	KpI	ST478	R	R	R	R	S	R
D30014	2005	Blood	Child	KpI	ST560	R	R	R	S	S	R
BKQ79K_2	2013	Blood	Child	KpI	ST609	S	S	S	S	S	S
7776	1998	Blood		KpI	ST664	R	S	S	S	S	S
D36068	2006	Blood	Child	KpI	ST664	R	R	R	S	S	R
D35263	2005	Blood	Child	KpI	ST664	R	R	R	S	S	R
D40443	2006	Blood	Child	KpI	ST664	R	R	R	S	R	R
BKQ2QU	2013	Blood	Child	KpI	ST874						
B10140	2005	CSF	Adult	KpI	ST881	R	S	S	S	S	S
1027734	2012	Blood	Adult	KpI	ST896	R	R	R	R	R	S
BHD3XR	2014	Blood	Adult	KpIII	ST1023						
D42680	2007	Blood	Child	KpI	ST1128	R	R	R	S	R	R
B1874	1999	Blood	Child	KpIII	ST1139	R	S	S			S
8407	1998	Blood		KpI	ST1425	R	R	R	S	S	S
C35	2009	RS	Adult	KpI	ST1916	R	R	R	R	S	R
A13	2009	RS	Adult	KpII		R	R	R	R	S	R
A44754	2007	Blood	Adult	KpI		R	R	R	R	S	R
D47091	2008	Blood	Child	KpI		R	R	R	R	S	R
C28	2009	RS	Adult	KpI		R	S	R	R	S	R
A28	2009	RS	Adult	KpI	ST14	R	R	R	R	S	R
C32	2009	RS	Adult	KpI		R	R	R	R	S	R
6	1996	Blood	Adult	KpI		R	S	S	S	S	S
C23	2009	RS	Adult	KpII		R	R	R	R	R	S
10038396	2010	Blood	Adult	KpI		R	R	R	R	S	R

Appendix 7 List of all AMR genes identified in Malawian *K. pneumoniae* genomes

AMR gene	Description	Resistance	Prevalence	
			n	%
fosA	metalloglutathione transferase	Fosmycin	69	97.2%
oqxA	Oqx Efflux pump gene	Fluoroquinolones	67	94.4%
oqxB	Efflux pump gene	Fluoroquinolones	66	93.0%
Sul2	Sulfonamide-resistant dihydropteroate synthase	Sulphanomides/cotrimoxazole	56	78.9%
dfrA	Dihydrofolate reductase	Methaxazole	58	81.7%
aac(6')	Acetyltransferase	Aminoglycoside, Fluoroquinolone	53	74.6%
$bla_{\text{TEM-1}}$	Beta-lactamase	Aminopenicillins	53	74.6%
catA	acetyltransferase	Chloramphenicol	53	74.6%
strB	Streptomycin phosphotransferase	Aminoglycosides	46	64.8%
strA	Streptomycin phosphotransferase	Aminoglycosides	46	64.8%
$bla_{\text{CTX-M-15}}$	ESBL	Beta-lactams	28	39.4%
Sul1	Sulfonamide-resistant dihydropteroate synthase	Sulphanomides/cotrimoxazole	25	35.2%
$bla_{\text{SHV-1}}$	Beta-lactamase	Beta-lactams	22	31.0%
tetD	Tetracycline efflux gene	Tetracyclines	21	29.6%
$bla_{\text{SHV-11}}$	Beta-lactamase	Beta-lactams	17	23.9%
mphA	Macrolide phosphotransferase	Chloramphenicol	16	22.5%
tetA	Tetracycline efflux gene	Tetracyclines	12	16.9%
$bla_{\text{OXA-1}}$	Beta-lactamase gene	Beta-lactams	11	15.5%
aadA2	Tetracycline efflux gene	Aminoglycosides	11	15.5%
$bla_{\text{SHV-28}}$	ESBL	Beta-lactams	10	14.1%
arr	ADP-ribosylation catalysing enzyme gene	Rifampin	7	9.9%
alph3	Aminoglycoside phosphotransferase	Aminoglycosides	6	8.5%
qnrB	PMQR gene	Fluoroquinolone	6	8.5%
cmlA1	MFS transporter/ chloramphenicol efflux gene	Chloramphenicol	5	7.0%
floR	Transmembrane segments efflux gene	Chloramphenicol/Florfenicol	5	7.0%
$bla_{\text{OXA-10}}$	ESBL gene	Beta-lactam	5	7.0%
tetB	Tetracycline efflux gene	Tetracycline	5	7.0%
qnrs	PMQR gene	Quinolones	4	5.5%
$bla_{\text{SCO-1}}$	Beta-lactamase gene	Aminopenicillins	4	5.5%

bla$_{\text{OXA-9}}$	ESBL gene	Aminopenicillins, cephalosporins	4	5.5%
bla$_{\text{SHV-12}}$	ESBL gene	Aminopenicillins, cephalosporins	3	4.2%
bla$_{\text{SHV-26}}$	ESBL gene	Aminopenicillins, cephalosporins	3	4.2%
bla$_{\text{OKPB}}$	Beta-lactamase gene	Aminopenicillins, cephalosporins	2	2.8%
ereA	Erythromycin esterase	Erythromycin	2	2.8%
bla$_{\text{SHV-7}}$	ESBL gene	Aminopenicillins	2	2.8%
bla$_{\text{TEM-63}}$	ESBL gene	Aminopenicillins, cephalosporins	1	1.4
bla$_{\text{LEN-16}}$	Beta-lactamase gene	Aminopenicillins	1	1.4%
bla$_{\text{LEN-25}}$	ESBL gene	Aminopenicillins, cephalosporins	1	1.4%
bla$_{\text{SHV-133}}$	ESBL gene	Aminopenicillins, cephalosporins	1	1.4%
bla$_{\text{SHV-25}}$	ESBL gene	Aminopenicillins, cephalosporins	1	1.4%
bla$_{\text{SHV-27}}$	ESBL gene	Aminopenicillins, cephalosporins	1	1.4%
bla$_{\text{SHV-36}}$	ESBL gene	Aminopenicillins, cephalosporins	1	1.4%
bla$_{\text{SHV-37}}$	ESBL gene	Aminopenicillins, cephalosporins	1	1.4%

Appendix 8 Genes identified in Gubbins predicted recombinant regions of Malawian *K. pneumoniae* KpI genome sequences (Feature Start and Feature End coordinates are based on the genome sequence of the reference strain *K. pneumoniae* NTUH2044)

Gene Name	Product Name	Feature Start	Feature End
abgB	aminobenzoyl-glutamate utilization protein	2447668	2449114
abgR	putative LysR-family transcriptional regulator	2450590	2451499
abgT	aminobenzoyl-glutamate transport protein	2446022	2447555
accC	acetyl-CoA carboxylase	4743134	4744484
aceB	malate synthase	264584	266186
aceK	bifunctional isocitrate dehydrogenase kinase/phosphatase protein	267672	269457
acnA	aconitate hydratase	2224309	2226982
acnB	aconitate hydratase	932597	935222
acoA	acetoin:DCPIP oxidoreductase alpha subunit	1092277	1093237
acoB	acetoin:DCPIP oxidoreductase beta subunit	1091246	1092275
acoC	dihydrolipoamide acetyltransferase	1089707	1091243
acoD	dihydrolipoamide dehydrogenase	1088320	1089718
acoK	trans-acting regulatory protein of aco operon	1093492	1096258
acrB	acriflavine resistance protein B	1269448	1272595
acrE	transmembrane protein affecting septum formation and cell membrane permeability	4749931	4751071
acrF	integral transmembrane protein of acridine resistance	4751083	4754194
ada	O6-methylguanine-DNA methyltransferase/DNA-binding transcriptional dual regulator	3696473	3697538
ade	putataive amidohydrolase	3604788	3606642
adhP	alcohol dehydrogenase	2781013	2782024
aes	acetyl esterase	41624	42542
ahpF	alkyl hydroperoxide reductase FAD/NAD(P)-binding subunit	1548302	1549868
aidB	putative acyl coenzyme A dehydrogenase	456305	457961
aldB	aldehyde dehydrogenase B	4086428	4087949
alkA	3-methyl-adenine DNA glycosylase II	3552858	3553707
alr	alanine racemase	308342	309422
amiC	N-acetylmuramoyl-L-alanine amidase	4305668	4306958
amn	AMP nucleosidase	3385240	3386695
ampE	putative transmembrane protease	916503	917358
ampG	beta-lactamase synthesis regulator/muropeptide transporter	1220301	1221777

amtB	probable ammonium transporter	1242660	1243947
ansA	cytoplasmic asparaginase I	2145380	2146400
apaH	diadenosinetetraphosphatase	836105	836954
aphA	diadenosine tetraphosphatase	310961	311696
araA	L-arabinose isomerase	855872	857375
araB	ribulokinase	857385	859095
arcB	aerobic respiration sensor-response protein	4689783	4692123
ardC	antirestriction protein	3463743	3464691
argG	argininosuccinate synthase	4658619	4659963
argS	arginyl-tRNA synthetase	3334982	3336716
arnT	4-amino-4-deoxy-L-arabinose transferase	4912205	4913861
aroD	3-dehydroquinate dehydratase	870561	871329
aroL	shikimate kinase II	1156100	1156634
arsB	Ars family arsenical pump	4397261	4398554
arsC	arsenate reductase	4396829	4397252
artM	arginine 3rd transport system permease component	1798811	1799480
asmA	suppressor of ompF assembly mutants	3549552	3551400
asnA	asparagine synthetase	1429	2422
asnS	asparaginyl-tRNA synthetase	1881123	1882524
aspS	aspartyl-tRNA synthetase	3327607	3329395
astE	succinylglutamate desuccinylase	2170861	2171827
baeR	two-compoent regulatory system response regulator	3571339	3572062
baeS	two-compoent regulatory system sensor protein	3569864	3571343
betA	choline dehydrogenase	1467409	1469119
betB	NAD+-dependent betaine aldehyde dehydrogenase	1469087	1470560
betT	high-affinity choline transport protein	1471289	1473323
bglA	6-phospho-beta-glucosidase A	4419966	4421400
bglX	beta-D-glucoside glucohydrolase	3620338	3622636
bioB	biotin synthetase	1693060	1694101
bipA	GTP-binding elongation factor family protein	36778	38602
birA	biotin--protein ligase	218911	219874
btuB	vitamin B12/cobalamin outer membrane transporter	117390	119229
btuC	vitamin B12 ABC transport system permease component	3105573	3106560
cadB	lysine/cadaverine transport protein	1354908	1356243
carA	carbamoyl-phosphate synthase small subunit	827682	828858
carB	carbamoyl-phosphate synthase large subunit	828875	832100
cas1	CRISPR-associated Cas1 family protein	3002915	3003800
cbiQ	putative cobalt transport protein	4273023	4273701
cca	tRNA nucleotidyl transferase	4535589	4536831
ccmF	cytochrome c-type biogenesis protein	2989590	2991543
ccmH	possible subunit of heme lyase	2988581	2989040

ccmI	possible subunit of heme lyase	2987742	2988585
celB	PEP-dependent phosphotransferase enzyme II	2175902	2177261
celF	6-phospho-beta-glucosidase	517920	519228
cfa	cyclopropane-fatty-acyl-phospholipid synthase	2927859	2929008
citA	putative citrate utilization protein A	1601076	1601592
citX	phosphoribosyl-dephospho-CoA transferase	848078	848666
citY	two-component sensor kinase for citrate	844632	846291
citZ	two-component response regulator for citrate	843944	844646
clpB	ATP-dependent protease	4005959	4008545
cls	cardiolipin synthetase	3125950	3127411
cobD	threonine-phosphate decarboxylase	4304574	4305651
cobT	nicotinate-nucleotide dimethylbenzimidazole-P phophoribosyl transferase	4269032	4270088
codB	probable transporter	2238296	2239742
cpdB	2',3'-cyclic nucleotide 2'-phosphodiesterase/3'-nucleotidase bifunctional periplasmic precursor protein	477286	479230
cpxA	sensor protein of stress-related two-component regulatory system	76798	78172
creB	catabolic regulation response regulator	791926	792616
csiE	stationary phase inducible protein	3951010	3952288
cstA	carbon starvation protein	1507516	1509622
cyaA	adenylate cyclase	169143	171696
cycA	D-alanine/D-serine/glycine transport protein	471682	473086
cydA	cytochrome d terminal oxidase polypeptide subunit I	1637772	1639344
cydC	cytochrome-related transport system ATP-binding component	1825863	1827585
cydD	cytochrome-related transport system ATP-binding component	1827585	1829352
cysA	ATP-binding component of sulfate permease A protein	3842662	3843757
cysD	sulfate adenylyltransferase subunit 2	4203409	4204318
cysE	serine acetyltransferase	5047439	5048261
cysH	phosphoadenosine phosphosulfate reductase	4207100	4207835
cysM	cysteine synthase B/O-acetylserine sulfhydrolase B	3841632	3842544
cysN	sulfate adenylyltransferase subunit 1	4201972	4203400
cysP	sulfate/thiosulfate transport system thiosulfate-binding component	3845454	3846471
cysS	cysteinyl-tRNA synthetase	1335266	1336652
dalT	D-arabinitol transporter	3582581	3583859
damX	putative membrane protein	4819677	4820964
dapB	dihydrodipicolinate reductase	826430	827252
dbpA	ATP-dependent RNA helicase	2352539	2353913
dcm	DNA cytosine methylase	3373381	3374815
dcp	dipeptidyl carboxypeptidase II	2507524	2509570
dcuS	two-component system sensor kinase for C4-dicarboxylate regulation of anaerobic fumurate respiratory system	2468922	2470611
ddl	D-alanylalanine synthetase	899386	900307

dedA	putative integral membrane protein	3780751	3781411
def	peptide deformylase	4765137	4765647
degQ	serine endoprotease	4708885	4710253
deoB	phosphopentomutase	777246	778470
dgoA	putative keto-hydroxyglutarate-aldolase/keto-deoxy-phosphogluconate aldolase	5202141	5202759
dgoD	putative mandelate racemase/muconate lactonizing enzyme	5200996	5202145
dgoK	2-oxo-3-deoxygalactonate kinase	5202742	5203621
dhaB1	glycerol dehydratase	4563389	4565057
dhaF	glycerol dehydratase-reactivation factor	4560528	4562352
dhaK	dihydroxyacetone kinase subunit K	4571701	4572772
dhaR	glycerol metabolism operon regulatory protein	4567966	4569892
dhaT	1,3-propanediol oxidoreductase	4565920	4567084
dinB	DNA polymerase IV	1065967	1067023
dinF	DNA-damage-inducible protein F	300931	302287
dinG	LexA-regulated (SOS) repair enzyme	1725484	1727641
dkgB	2,5-diketo-D-gluconate reductase B	1041531	1042335
dmsA	anaerobic dimethyl sulfoxide reductase subunit A	1839502	1841941
dnaA	chromosomal replication initiation protein	5212741	5214145
dnaB	replicative DNA helicase	306895	308311
dnaE	DNA polymerase III subunit alpha	1017535	1021018
dnaJ	chaperone protein	813194	814328
dnaX	DNA polymerase III subunits gamma and tau	1281050	1282958
dppA	putative dipeptide ABC transport system periplasmic binding component	3578676	3580398
dppC	putative dipeptide ABC transport system permease component	3576740	3577646
dppF	dipeptide transport system ATP-binding component	4984484	4985498
dppF2	putative dipeptide ABC transport system ATP-binding component	3575121	3575895
dsbB	disulfide bond formation protein B	3260979	3261510
dsbC	protein disulfide isomerase II	4413689	4414403
dsbD	thiol:disulfide interchange protein precursor	400032	401829
dsbG	periplasmic disulfide isomerase/thiol-disulphide oxidase	1546390	1547140
dsdA	D-serine deaminase	5187288	5188617
dsdC	putative LysR-family transcriptional regulator	5184785	5185709
eco	ecotin precursor	3688890	3689334
eda	keto-hydroxyglutarate-aldolase/keto-deoxy-phosphogluconate aldolase	3312054	3312696
efp	elongation factor P	422414	422981
ego	Putative ABC transport system ATP-binding component	4584203	4585739
eitD	putative transport system membrane component	3085915	3087238
emrA	multidrug resistance secretion protein	4108111	4109284
emrD	multidrug MFS integral membrane transporter	5183441	5184626

endA	DNA-specific endonuclease I	4457562	4458327
entB	2,3-dihydro-2,3-dihydroxybenzoate synthetase, isochroismatase	1505139	1505991
entC	isochorismate synthase	1502321	1503509
entD	enterochelin synthetase component D	1488319	1488949
entE	2,3-dihydroxybenzoate-AMP ligase	1503518	1505126
entF	ATP-dependent serine activating enzyme	1492953	1496835
epd	D-erythrose 4-phosphate dehydrogenase	4445140	4446169
eutA	putative chaperonin in ethanolamine utilization	3856186	3857590
eutB	ethanolamine ammonia-lyase large subunit	1130283	1131672
eutE	ethanolamine utilization protein	3861044	3862448
eutG	putative transport protein in ethanolamine utilization	3859018	3860206
eutP	putative ethanolamine utilization protein	3865680	3866160
fabF	3-oxoacyl-(acyl carrier protein) synthase	2029883	2031125
fabZ	putative (3R)-hydroxymyristoyl-ACP dehydratase	1014451	1014979
fadA	acetyl-CoA acetyltransferase	204295	205459
fadB	3-hydroxyacyl-CoA dehydrogenase/3-hydroxybutyryl-CoA epimerase/delta(3)-cis-delta(2)-trans-enoyl-CoA isomerase/enoyl-CoA hydratase	205468	207694
fadH	2,4-dieonyl-CoA reductase	4588700	4590722
fadR	fatty acid metabolism regulator	3258332	3259127
fbp	fructose-1,6-bisphosphatase	495682	496681
fdhF	formate dehydrogenase-H	356721	358869
fdnG	formate dehydrogenase-N alpha subunit	2794352	2797400
fdnH	formate dehydrogenase-N beta subunit	2793456	2794341
fdnI	formate dehydrogenase-N gamma subunit	2792807	2793464
fdoG	formate dehydrogenase-O alpha subunit	46018	49069
feaR	2-phenylethylamine catabolism regulatory protein	2367239	2368202
fecB	iron(III) dicitrate-binding periplasmic protein	3083197	3084196
feoB	ferrous iron transport protein B	4841999	4844318
fepA	ferrienterobactin receptor	318495	320754
fepG	ferric enterobactin transport protein	1497690	1498683
fes	enterochelin esterase	1491503	1492712
ffh	SRP component with 4.5S RNA	4023234	4024599
fhlA	formate hydrogen-lyase transcriptional activator	4158466	4160539
fhuF	ferric hydroxamate transport protein	762715	763504
fim	probable pilin chaperone	4076316	4076985
fimA	type 1 fimbrial protein	4075723	4076275
fimD	outer membrane protein for export and assembly of type 1 fimbriae	4077014	4079552
fimE	site-specific recombinase involved in flagellar phase switching/type 1 fimbriae regulatory protein with FimB	4379511	4380120
fimF	type 1 fimbrial minor component	4385217	4385748

fimG	type 1 fimbrial minor component	4385760	4386261
fimZ	putative transcriptional regulator of fimbrial expression	4074918	4075620
fklB	FKBP-type peptidyl-prolyl cis-trans isomerase	470697	471378
fldB	flavodoxin	4415422	4415944
fmt	methionyl-tRNA formyltransferase	4765661	4766609
folC	dihydrofolate:folylpolyglutamate synthetase/dihydrofolate synthetase	3778301	3779570
folD	5,10-methylene-tetrahydrofolate dehydrogenase/cyclohydrolase	1338833	1339700
frdA	fumarate reductase	426099	427890
fre	NAD(P)H-flavin reductase	203487	204189
frnA	ABC-type transporter periplasmic component	2782193	2783816
frnB	ABC-type transporter permease component	2783826	2784771
frnD	ABC-type transporter ATP-binding component	2785594	2787229
fruA	fructose-specific PTS family enzyme IIBBC component 3655849 3657544	910632	911676
gudP	permease of the major facilitator superfamily	868085	869396
gutQ	putative polysialic acid capsule expression protein	4132685	4133651
gyrA	DNA gyrase subunit A	3706977	3709611
hcaE	3-phenylpropionate dioxygenase alpha subunit	2800794	2801886
hcaF	3-phenylpropionate dioxygenase beta subunit	2800309	2800795
hcr	NADH oxidoreductase for HCP	1811411	1812380
hemF	coproporphyrinogen III oxidase	3850686	3851586
hepA	ATP-dependent helicase	849412	852319
hflC	protease specific for phage lambda cII repressor	449142	450147
hipB	murein or DNA biosynthesis regulatory protein	2590807	2591095
hisA	N-(5-phospho-L-ribosyl-formimino)-5-amino-1-(5-phosphoribosyl)-4-imidazolecarboxamide isomerase	3505496	3506237
hisB	imidazoleglycerolphosphate dehydratase/histidinol-phosphate phosphatase	3503842	3504910
hisC	histidinol-phosphate aminotransferase	3502784	3503846
hisD	histidinol dehydrogenase	3501483	3502788
hisF	imidazole glycerol phosphate synthase catalytic subunit	3506218	3506995
hisH	imidazole glycerol phosphate synthase glutamine amidotransferase subunit	3504909	3505500
hisI	bifunctional phosphoribosyl-AMP cyclohydrolase/phosphoribosyl-ATP pyrophosphatase protein	3506988	3507588
hisM	histidine transport system permease component	3771082	3771799
hisP	histidine transport system ATP-binding component	3770301	3771075
hisQ	histidine transport system permease component	3771795	3772482
hmsS	haemin storage system HmsS protein	388075	388528
hmuU	hemin transport system permease component	4183935	4184931
hofB	protein transport protein	912907	914293
holB	DNA polymerase III subunit delta	2033711	2034716

holC	DNA polymerase III subunit chi	536788	537265
hpaB	4-hydroxyphenylacetate 3-hydroxylase	721104	722667
hpaC	4-hydroxyphenylacetate 3-monooxygenase coupling protein	720572	721085
hpaD	3,4-dihydroxyphenylacetate 2,3-dioxygenase	727282	728140
hpaE	5-carboxymethyl-2-hydroxymuconate semialdehyde dehydrogenase	728141	729608
hpaF	5-carboxymethyl-2-hydroxy-muconic acid isomerase	726892	727273
hpaG1	4-hydroxyphenylacetate degradation bifunctional isomerase/decarboxylase N-terminal subunit	730365	730998
hpaG2	4-hydroxyphenylacetate degradation bifunctional isomerase/decarboxylase C-terminal subunit	729604	730369
hpaH	2-oxo-hepta-3-ene-1,7-dioic acid hydratase	725980	726784
hpaI	2,4-dihydroxyhept-2-ene-1,7-dioic acid aldolase	725172	725970
hpaR	homoprotocatechuate degradative operon repressor	731285	731726
hpaX	4-hydroxyphenylacetate permease	723792	725163
hrpA	ATP-dependent helicase	2389433	2393336
htpG	heat shock protein 90	1284040	1285939
htrB	lipid A biosynthesis lauroyl acyltransferase	2005042	2005963
hutC	histidine utilization repressor C	1684672	1685398
hutH	histidine ammonia lyase	1687225	1688752
hutI	imidazolonepropionase	1682272	1683658
hycE	hydrogenase 3 large subunit	4148793	4150503
hydN	electron transport protein	5032658	5033144
hypD	hydrogenase expression/formation protein	4156101	4157223
hypF	hydrogenase maturation protein	4139036	4141271
hyuE	hydantoin racemase	2704150	2704894
icdA	isocitrate dehydrogenase	2088248	2089499
idi	isopentenyl-diphosphate delta-isomerase	4408450	4409005
ileS	isoleucyl-tRNA synthetase	818326	821143
ilvB	acetolactate synthase large subunit	5179186	5180875
ilvD	dihydroxy-acid dehydratase	132208	134059
ilvE	branched-chain amino acid aminotransferase	131218	132148
ilvG	acetolactate synthase II large subunit	129301	130948
ilvM	acetolactate synthase II small subunit	130944	131202
int	integrase	3395996	3397259
iolD	acetolactate synthase	551844	553785
iolE	sugar phosphate isomerases/epimerases	555794	556694
iroB	putative UDP-glucoronosyl and UDP-glucosyl transferase	3436633	3437749
iroN	TonB dependent outer membrane siderophore receptor protein	2188601	2190848
irp1	yersiniabactin biosynthetic protein	3411160	3420652
ispA	geranyltranstransferase	1198124	1199024
ispH	4-hydroxy-3-methylbut-2-enyl diphosphate reductase	822164	823160
kbl	2-amino-3-ketobutyrate coenzyme A ligase	5056691	5057885

kdgK	ketodeoxygluconokinase	4951971	4952901
kdpA	potassium-transporting ATPase subunit A	1613163	1614987
kdpD	sensor for high-affinity potassium transport system	1607820	1610508
kduD	2-deoxy-D-gluconate 3-dehydrogenase	4357082	4357844
kefA	putative small-conductance mechanosensitive channel	1274733	1278081
kefB	glutathione-regulated potassium-efflux system protein	4795197	4797003
kefC	glutathione-regulated potassium-efflux system protein	833160	835026
lacZ	beta-D-galactosidase	2528561	2531669
lamB	maltoporin precursor	293309	294599
lasT	putative tRNA/rRNA methyltransferase	796870	797557
lexA	LexA repressor	300292	300901
livF	ATP-binding component of leucine transport	3225602	3226322
livG	high-affinity branched-chain amino acid transporter ATP-binding component	4887473	4888241
livH	ABC transport system permease component	1126909	1127809
livJ	high-affinity branched-chain amino acid transporter periplasmic binding component	4892684	4893803
livM	ABC transport system permease component	1125860	1126910
lpdA	dihydrolipoamide dehydrogenase	927844	929272
lpxD	UDP-3-O-(3-hydroxymyristoyl)-glucosamine N-acyltransferase	283119	283725
ltrA	probable HTH-type transcriptional regulator	824484	825414
lysR	positive regulator for lys	4343160	4344087
lysS	lysyl-tRNA synthetase	4409239	4410757
lyx	putative L-xylulose kinase	1084460	1086026
maeB	phosphate acetyltransferase	3866726	3869006
malI	maltose regulon regulatory protein	2424000	2425029
malM	periplasmic protein of mal regulon	294713	295637
malX	maltose/glucose-specific PTS family enzyme IIBC component	2422235	2423828
malZ	maltodextrin glucosidase	1170125	1171943
manB	phosphomannomutase	3522915	3524286
manC	mannose-1-phosphate guanylyltransferase	3524308	3525724
mdaB	modulator of drug activity B	4516665	4516890
mdoB	phosphoglycerol transferase I	750808	753100
melA	alpha-galactosidase	379495	380851
menC	O-succinylbenzoate synthase	3731261	3732227
menD	2-oxoglutarate decarboxylase	3733853	3735542
menE	O-succinylbenzoic acid--CoA ligase	3729888	3731265
mepA	penicillin-insensitive murein endopeptidase	3792121	3792946
metB	putative cystathionine gamma-synthase	4484446	4485595
metG	methionyl-tRNA synthetase	3611064	3613107
metH	B12-dependent methionine synthase	270442	274216
metL	bifunctional aspartate kinase II/homoserine dehydrogenase II	92086	94519

mexF	integral transmembrane protein for acridine resistance	4037510	4040663
mglC	methyl-galactoside transport system permease component	3640327	3641338
mgsA	methylglyoxal synthase	1919141	1919729
mgtA	P-type Mg2+ transport ATPase	525245	527954
mhpA	3-(3-hydroxyphenyl)propionate hydroxylase	3052060	3053725
mhpB	2,3-dihydroxyphenylpropionate 1,2-dioxygenase	3051114	3052059
mhpC	2-hydroxy-6-ketonona-2,4-dienedioic acid hydrolase	3050230	3051160
mhpD	2-keto-4-pentenoate hydratase	3049413	3050223
mhpE	4-hydroxy-2-ketovalerate aldolase	3047457	3048474
mhpF	acetaldehyde dehydrogenase	3048470	3049421
mhpT	putative transport protein	3046185	3047382
miaE	hydroxylase for synthesis of 2-methylthio-cis-ribozeatin in tRNA	532475	533240
minE	cell division topological specificity factor	3265174	3265444
mltC	membrane-bound lytic murein transglycosylase C	4473507	4474593
mntR	putative Mn-dependent transcriptional regulator	1741318	1741792
mobB	mobilization protein	3460902	3462792
mocB	putative rhizopine uptake ABC transport system periplasmic solute-binding protein precursor	3189389	3190304
modB	molybdenum transport system permease component	1662329	1663019
modF	molybdate transport system ATP-binding component	1658783	1660268
mogA	molybdenum cofactor biosynthesis protein	806879	807482
mpl	UDP-N-acetylmuramate:L-alanyl-gamma-D-glutamyl-meso-diaminopimelate ligase	496856	498230
mppA	putative transport periplasmic protein	2265373	2266990
mrcB	penicillin-binding protein 1b	978212	980765
mrdB	rod shape-determining membrane protein	1562497	1563610
mreC	rod shape-determining protein	4735496	4736489
mrkC	fimbrial biogenesis outer membrane usher protein mrkC precursor	4369872	4372359
mtlA	mannitol-specific PTS family enzyme II component	5038296	5040204
mtnC	2,3-diketo-5-methylthio-1-phosphopentane phosphatase	1541355	1542114
mtnK	5-methylthioribose kinase	1517977	1519177
mukB	condesin subunit B	1870051	1874500
mukF	condesin subunit F	1868044	1869367
murD	UDP-N-acetylmuramoyl-L-alanyl-D-glutamate synthetase	894213	895530
mutS	DNA mismatch repair protein	4185856	4188418
mviN	putative virulence factor	2015059	2016634
nadB	L-aspartate oxidase	3986342	3987962
nadE	NAD(+) synthetase	2173707	2174535
nagC	transcriptional repressor of genes for catabolic enzyme	1591399	1592620
narG	nitrate reductase 1 alpha subunit	3150810	3154554
narH	nitrate reductase 1 beta subunit	3149278	3150814

narI	nitrate reductase 1 gamma subunit	3147894	3148572
narK	nitrite extrusion protein	3154959	3156348
narV	cryptic nitrate reductase 2 gamma subunit	2815596	2816277
narW	cryptic nitrate reductase 2 delta subunit	2814904	2815600
narY	cryptic nitrate reductase 2 beta subunit	2813360	2814905
narZ	cryptic nitrate reductase 2 alpha subunit	2809623	2813364
ndk	nucleoside diphosphate kinase	3928466	3928898
nei	endonuclease VIII	1624333	1625125
nfi	endonuclease V	248116	248788
nfsA	oxygen insensitive NADPH nitroreductase	1786959	1787682
nhaB	pH independent Na+/H+ antiporter	3259081	3260851
nhaR	LysR family transcriptional activator of cation transport	815735	816632
nhoA	N-hydroxyarylamine O-acetyltransferase	2816946	2817792
nikA	nickel transport system periplasmic binding component	4920865	4922434
nikB	nickel transport system permease component	4922433	4923378
nikC	nickel transport system permease component	4923374	4924208
nirB	nitrite reductase large subunit	4811859	4814403
nlpC	lipoprotein	1106109	1106676
norA	multidrug efflux protein	2929909	2931283
norV	anaerobic nitric oxide reductase flavorubredoxin	4135351	4136800
nrdA	ribonucleotide-diphosphate reductase alpha subunit	3710830	3713116
nrdE	ribonucleotide-diphosphate reductase alpha subunit	4099318	4101460
nrdG	anaerobic ribonucleotide reductase activating protein	513229	513694
nuoC	NADH dehydrogenase I chain CD	3750518	3752327
nuoH	NADH dehydrogenase subunit H	3744948	3745926
nuoN	NADH dehydrogenase subunit N	3738525	3739983
nupG	nucleoside transport protein	4474784	4476038
oadA	oxaloacetate decarboxylase alpha chain	4713510	4715283
oadB	oxaloacetate decarboxylase beta chain	4712193	4713495
ompC	outer membrane porin protein C	3698767	3699865
ompF	outer membrane protein 1A/OmpK35 porin	1879452	1880532
ompN	non-specific outer membrane pore protein N	2355999	2357124
ompS	outer membrane pore protein S1	3376428	3377574
oppB	oligopeptide ABC transport system permease component	3131670	3132591
otsA	trehalose-6-phosphate synthase	3340309	3341734
paaA	phenylacetic acid degradation protein	2376997	2377927
paaB	phenylacetic acid degradation protein	2377938	2378226
paaC	phenylacetic acid degradation protein	2378233	2378989
paaD	phenylacetic acid degradation protein	2378982	2379498
paaE	probable phenylacetic acid degradation NADH oxidoreductase	2379505	2380576
paaF	probable enoyl-CoA hydratase	2380572	2381340

paaG	enoyl-CoA hydratase	2381342	2382131
paaI	phenylacetic acid degradation protein	2383545	2383968
paaK	phenylacetate-CoA ligase	2385199	2386516
paaX	transcriptional repressor for phenylacetic acid degradation	2386635	2387562
paaZ	aldehyde dehydrogenase/enoyl-CoA hydratase	2374665	2376711
pabB	para-aminobenzoate synthase component I	3272662	3274087
panC	pantoate--beta-alanine ligase	950644	951499
parE	DNA topoisomerase IV subunit B	4517359	4519255
pbpC	penicillin binding protein 1C	3929058	3931383
pbpG	D-alanyl-D-alanine endopeptidase	3625557	3626505
pckA	phosphoenolpyruvate carboxykinase	4834428	4836051
pcnB	poly(A) polymerase I	952898	954260
pduC	propanediol dehydratase large subunit	4288914	4290579
pduN	propanediol utilization protein	4295889	4296165
pduT	propanediol utilization polyhedral body protein	4301052	4301607
pduU	propanediol utilization polyhedral body protein	4301606	4301957
pduW	acetate/propionate kinase	4302398	4303613
pduX	propanediol utilization protein	4303668	4304571
pdxA	4-hydroxythreonine-4-phosphate dehydrogenase	2555518	2556505
pdxB	erythronate-4-phosphate dehydrogenase	3783326	3784463
pdxK	pyridoxine kinase	3838852	3839719
pdxY	pyridoxine kinase	2903797	2904658
pepA	leucyl aminopeptidase	537351	538863
pepB	aminopeptidase B	3939410	3940727
pepE	peptidase E	276089	276779
pepN	aminopeptidase N	1890115	1892731
pepP	proline aminopeptidase P II	4433374	4434691
pepQ	proline dipeptidase	207849	209181
pflB	formate acetyltransferase 1	1845951	1848234
pgi	glucose-6-phosphate isomerase	281420	283070
pgtB	phosphoglycerate transport system sensor protein	2637769	2639779
pgtP	phosphoglycerate transporter protein	2641401	2642793
pheA	bifunctional chorismate mutase P/prephenate dehydratase	4012100	4013261
phnD	phosphonate ABC transport system periplasmic phosphonate-binding component	1111273	1112200
phnH	carbon-phosphorus lyase complex subunit	373310	373895
phnI	bacterial phosphonate metabolism protein	372245	373310
phnM	phosphonate metabolism protein	368728	369865
phnP	phosphonate utilization protein	367391	368150
phnS	2-aminoethylphosphonate transport system periplasmic binding component	5167013	5168027
phnT	2-aminoethylphosphonate transporter ATP-binding component	5165895	5167008

phnU	2-aminoethylphosphonate transport system memrane component	5165035	5165896
phnV	2-aminoethylphosphonate transport system memrane component	5164235	5165033
phnW	2-aminoethylphosphonate--pyruvate transaminase	5169066	5170170
phnX	phosphonoacetaldehyde hydrolase	5170179	5170989
phoE	outer membrane pore protein E precursor	1105012	1106065
phoP	response regulator in two-component regulatory system with PhoQ	2081388	2082102
phoQ	sensor kinase in two-component regulatory system with PhoP	2079922	2081389
phoU	high-affinity phosphate transport system negative regulator for pho regulon	5229262	5229988
pitA	putative low-affinity inorganic phosphate transporter	4929418	4930915
plsB	glycerol-3-phosphate acyltransferase	297249	299673
pmbA	putative antibiotic maturation protein	501865	503218
pntB	pyridine nucleotide transhydrogenase	2441152	2442541
polA	DNA polymerase I	25653	28446
polB	DNA polymerase II	852506	854864
potA	spermidine/putrescine transport ATP-binding protein	2073727	2074864
potI	spermidine/putrescine transport system permease component	1792943	1793831
ppc	phosphoenolpyruvate carboxylase	104670	107322
ppdA	prepilin peptidase dependent protein A	4322124	4322592
ppdB	prepilin peptidase dependent protein B	4321570	4322134
ppiA	peptidyl-prolyl cis-trans isomerase A	4808268	4808838
ppiB	peptidyl-prolyl cis-trans isomerase B	1334596	1335091
ppiD	peptidyl-prolyl cis-trans isomerase D	1230356	1232231
ppk	polyphosphate kinase	3901347	3903408
ppx	exopolyphosphatase	3903411	3904944
pqqF	pyrroloquinoline quinone synthesis protein F	2732501	2734787
priA	primosome assembly protein	87576	89844
priC	primosomal replication protein N	1279281	1279836
prkB	probable phosphoribulokinase	4801108	4801978
prlC	oligopeptidase A	4934742	4936785
prmA	ribosomal protein L11 methyltransferase	4746289	4747171
proA	gamma-glutamyl phosphate reductase	1102355	1103609
proB	gamma-glutamyl kinase	1103619	1104723
proY	putative amino acid permease	1688850	1690233
prpR	putative Fis-type transcriptional regulator	4088102	4090001
pspA	phage shock protein inner membrane protein	2253442	2254111
pspF	psp operon transcriptional activator	2252300	2253290
pstA	high-affinity phosphate transport system permease component	5230859	5231750
ptr	protease III	4314462	4317348
ptsI	PEP-protein PTS family enzyme I	3836296	3838024

pulA	pullulanase precursor	971641	974950
pulB	pullulanase-specific type II secretion system component B	976633	977212
pulC	pullulanase-specific type II secretion system component C	970492	971335
pulD	pullulanase-specific type II secretion system secretin	968509	970483
pulG	pullulanase-specific type II secretion system periplasmic pseudopilin	965360	965795
pulH	pullulanase-specific type II secretion system component H	964845	965361
pulN	pullulanase-specific type II secretion system component N	960513	961260
purD	phosphoribosylamine--glycine ligase	253990	255286
purE	phosphoribosylaminoimidazole carboxylase catalytic subunit	1332818	1333328
purH	bifunctional phosphoribosylaminoimidazolecarboxamide formyltransferase/IMP cyclohydrolase	255301	256891
purL	phosphoribosylformylglycinamidine synthase	3968277	3972165
pykA	pyruvate kinase	3317570	3319037
pykF	pyruvate kinase	3063991	3065404
pyrH	uridylate kinase	1004226	1004952
qor	quinone oxidoreductase	305742	306726
rbn	ribonuclease BN	39330	40191
rbtT	ribitol transporter	3591661	3592945
rcsC	hybrid sensory histidine kinase in two-component regulatory system with RcsB	3704005	3706846
rdgC	recombination associated protein	1158261	1159173
recA	recombinase A	4122689	4123748
recB	exonuclease V beta subunit	4310929	4314466
recD	exonuclease V alpha subunit	4309088	4310933
recG	ATP-dependent DNA helicase	5096132	5098214
recN	DNA repair protein	4028659	4030321
recR	recombination protein	1283347	1283953
relA	(p)ppGpp synthetase I	4218689	4220927
relB	putative helix-turn-helix protein	512709	512952
rep	ATP-dependent DNA helicase	139051	141073
rhaP	bacterial inner-membrane translocator	64509	65517
rhaQ	rhamnose ABC transport system permease component	63508	64513
rhlE	putative ATP-dependent RNA helicase	1723941	1725297
rho	transcription termination factor	147759	149091
ribH	riboflavin synthase beta subunit	1192714	1193185
rihC	nucleoside hydrolase	823226	824141
rluC	23S rRNA pseudouridylate synthase	2022918	2023872
rmlA	glucose-1-phosphate thymidylyltransferase	155023	155905
rncS	ribonuclease III	3979326	3979941
rnd	ribonuclease D	3265521	3266649
rne	RNase E	2018985	2022219
rnfB	electron transport complex protein	2895071	2895650

rnhA	RNase HI	1049346	1049925
rnr	ribonuclease R	452953	455386
rob	right origin-binding protein	790362	791232
rplY	50S ribosomal protein L25	3679920	3680205
rpoC	DNA-directed RNA polymerase beta subunit 231352 235576 rpoC KP1_0223 DNA-directed RNA polymerase beta subunit	231352	235576
rpoD	RNA polymerase sigma factor	4549609	4551508
rpoN	DNA-directed RNA polymerase subunit N	4684639	4686073
rpsA	30S ribosomal protein S1	1855722	1857396
rrmB	16S rRNA m5C967 methyltransferase	4766661	4767957
rseB	periplasmic negative regulator of sigma-E factor	3983707	3984664
rsmC	16S ribosomal RNA m2G1207 methyltransferase	765673	766702
sapB	ABC-type peptide transport system permease component	2242466	2243432
sbcB	exodeoxyribonuclease I	3492286	3493711
sbcD	ATP-dependent dsDNA exonuclease	1163386	1164592
sbmC	DNA gyrase inhibitor	3490292	3490766
scsB	copper-sensitivity suppressor protein B	4421933	4423943
scsC	copper-sensitivity suppressor protein C	4423944	4424556
sdhA	succinate dehydrogenase flavoprotein subunit	1627882	1629649
secA	preprotein translocase ATPase subunit	905512	908218
secF	preprotein translocase SecF subunit	1186953	1187925
selA	selenocysteine synthase	5035226	5036615
selB	selenocysteinyl-tRNA-specific translation factor	5033388	5035230
serA	D-3-phosphoglycerate dehydrogenase	4436792	4438025
serC	phosphoserine aminotransferase	1852285	1853374
sfcA	NAD-linked malate dehydrogenase	2774288	2775986
sfsA	sugar fermentation stimulation protein	955914	956631
sfuC	iron(III)-transport ATP-binding protein	1934590	1935619
sgaH	3-keto-L-gulonate 6-phosphate decarboxylase	464094	464745
shiA	shikimate transporter	3480863	3482183
sitA	iron transport system periplasmic binding component	4171444	4172362
sitC	iron transport system inner membrane permease component	4173176	4174025
slt	soluble lytic murein transglycosylase	786723	788685
smf	putative competence protein	4763884	4765009
smpA	small membrane protein A	4030468	4030810
smpB	SsrA tmRNA-binding protein	4031743	4032226
smtA	S-adenosylmethionine-dependent methyltransferase	1867235	1868048
sohB	putative peptidase	2217838	2218885
solA	putative sarcosine oxidase	2007403	2008522
soxS	regulation of superoxide response regulon	336260	336590
speA	arginine decarboxylase	4451195	4453184
speB	agmatinase	4450041	4450962

speC	ornithine decarboxylase isozyme	4476081	4478262
speG	spermidine N1-acetyltransferase	2500533	2501094
spoT	(p)ppGpp synthetase II	5093312	5095433
srlE	glucitol/sorbitol-specific PTS family enzyme IIBC component	4129187	4130168
srmB	ATP-dependent RNA helicase	3988815	3990147
ssb	single-strand DNA-binding protein	315680	316205
sseB	enhanced serine sensitivity	3938535	3939312
ssuB	putative aliphatic sulfonates transporter ATP-binding component	1892935	1893709
ssuD	putative alkanesulfonate monooxygenase	1894507	1895653
sucA	2-oxoglutarate decarboxylase	1630762	1633570
sufC	putative ATP-binding component of a transport system	3069994	3070741
sufD	Fe-S cluster assembly protein	3068745	3070020
tadA	tRNA-specific adenosine deaminase	3973953	3974496
talB	transaldolase	805841	806795
tauC	taurine transport system permease component	1137640	1138468
tauD	2-oxoglutarate-dependent taurine dioxygenase	1138464	1139316
thrC	threonine synthase	801316	802597
thrS	threonyl-tRNA synthetase	3116255	3118184
tkrA	2-hydroxyacid dehydrogenase	5006716	5007688
tktB	transketolase	3870234	3872268
togT	putative oligogalacturonide transporter	3347819	3349418
tolA	cell envelope integrity inner membrane protein	2654720	2655149
tolB	translocation protein TolB precursor	4358329	4359505
tolC	outer membrane channel precursor protein	4521954	4523430
trkA	potassium transporter peripheral membrane component	4767970	4769347
trmD	tRNA (guanine-N(1)-)-methyltransferase	4021360	4022128
trmU	tRNA (5-methylaminomethyl-2-thiouridylate)-methyltransferase	2085719	2086886
trpS	putative tryptophanyl-tRNA synthetase	692101	693109
trxC	putative thioredoxin-like protein	498869	500078
tynA	copper-requiring tyramine oxidase	2371921	2374189
tyrB	aspartate aminotransferase	309556	310792
tyrS	tyrosyl-tRNA synthetase	2904719	2905994
ubiA	4-hydroxybenzoate octaprenyltransferase	296334	297201
ubiB	putative ubiquinone biosynthesis protein	197119	198760
ubiD	3-octaprenyl-4-hydroxybenzoate decarboxylase	201962	203438
ubiF	2-octaprenyl-3-methyl-6-methoxy-1,4-benzoquinol hydroxylase	1586304	1587480
udk	uridine kinase	3552102	3552810
udp	uridine phosphorylase	193332	194094
ugd	UDP-glucose 6-dehydrogenase	3521585	3522752
uge	uridine diphosphate galacturonate 4-epimerase	3519636	3520641

ugpE	sn-glycerol 3-phosphate transport membrane component	4882583	4883429
uhpB	two-component regulatory system sensor histidine kinase	5176652	5178155
ulaA	ascorbate-specific PTS family enzyme IIC component	461885	463286
umuC	DNA polymerase V subunit C	2126901	2128167
ureD	urease accessory protein	4539368	4540193
uup	ABC transport system ATP-binding component	1902475	1904383
uvrB	excinuclease ABC subunit B	1698083	1700105
uxaA	altronate hydrolase	4598439	4599927
vagD	putative DNA binding protein	3432735	3433152
valS	valyl-tRNA synthetase	533933	536789
vanA	vanillate O-demethylase oxygenase subunit	2584843	2585881
virB1	type IV secretory pathway VirB1 component	3449224	3449935
virB11	type IV secretory pathway VirB11 component	3458116	3459142
virB2	type IV secretory pathway VirB2 component	3449934	3450228
virB3-4	type IV secretory pathway VirB3-4 component	3450240	3452979
virB5	type IV secretion system VirB5 component	3452996	3453701
wabG	glucuronic acid transferase	5066417	5067545
wabH	putative glycosyltransferase	5067547	5068642
wabN	deacetylase	5064270	5065293
wbbM	putative glycosyltransferase	3515171	3517067
wbbN	putative glycosyltransferase	3513111	3514005
wcaJ	probable CPS biosynthesis glycosyltransferase	3527529	3528933
wecE	TDP-4-oxo-6-deoxy-D-glucose transaminase	156562	157693
wecF	TDP-Fuc4NAc:lipidIIFuc4NAc transferase	158941	160018
wecG	UDP-N-acetyl-D-mannosaminuronic acid transferase	161372	162113
wza	putative capsule polysaccharide export protein precursor	3540982	3542116
wzb	putative protein tyrosine phosphatase	3540543	3540978
wzc	inner membrane tyrosine autokinase	3538376	3540530
wzi	outer membrane protein	3542258	3543773
wzm	lipopolysaccharide O-antigen ABC transport system transmembrane component	3517822	3518602
wzx	enterobacterial common antigen translocase	157694	158945
wzy	enterobacterial common antigen polymerase	160014	161361
xylH	D-xylose transport system permease component	5025391	5026573
xynB	puative beta-1,4-xylosidase	653955	655635
yaaH	putative integral membrane regulator	808968	809535
yabF	glutathione-regulated potassium-efflux system ancillary protein	832634	833168
yadF	putative carbonic anhdrase	945720	946383
yadQ	chloride channel protein	988516	989935
yaeF	putative synthase	1028545	1029388
yaeT	putative outer membrane antigen	1010352	1012782
yafC	putative LysR-family transcriptional regulator	1042380	1043286

yafM	conserved protein	2078086	2078578
yajF	possible NAGC-like transcriptional regulator	1159159	1160179
ybaR	copper-transporting P-type ATPase	1296057	1298559
ybaX	putative (aluminum) resistance protein	1233372	1234074
ybbF	UDP-2,3-diacylglucosamine hydrolase	1333870	1334593
ybbS	putative LysR-family transcriptional regulator	4361882	4362752
ybcO	putative benzoate membrane transport protein	2864495	2865662
ybdA	putative transport	1499799	1501041
ybdH	putative alcohol dehydrogenase	1511586	1512675
ybdJ	putative inner membrane protein	1432524	1432773
ybgR	putative CDF family transport protein	1650401	1651346
ybgS	putative homeobox protein	1651463	1651829
ybhC	putative pectinesterase	1680900	1682184
ybhE	putative isomerase	509544	510699
ybiI	hypothetical Zinc-finger containing protein	1729294	1729561
ybiL	putative outer membrane receptor for iron transport	2194551	2196837
ybiT	putative transport system ATP-binding component	1747455	1749048
ybiV	conserved protein with phophatase-like domain	1755636	1756452
ybiW	putative formate acetyltransferase 3	1756628	1759061
ybiX	putative hydroxylase	2193825	2194503
ybjE	putative surface protein	1814186	1815086
ybjJ	putative DeoR-type transcriptional regulator	1781854	1783057
ybjK	putative DeoR-type transcriptional regulator	1783139	1783703
ybtS	putative salicylate synthetase	3397452	3398757
ybtU	thiazolinyl-S-HMWP1 reductase	3420588	3421749
ycaI	putative recombination protein with metallo-hydrolase domain	1857881	1860305
ycaJ	putative polynucleotide enzyme	1836575	1837919
yccA	putative TEGT family carrier/transport protein	1924073	1924733
yccS	putative PET-family efflux transport protein	1914351	1916487
yccV	putative inner membrane protein	1920198	1920516
yccW	putative SAM-dependent methyltransferase	1920575	1921811
ycdH	putative flavin:NADH reductase	1969798	1970329
ycdT	putative transmembrane protein	763610	764657
yceF	putative inhibitor of septum formation	2023909	2024494
ycfD	putative enzyme	2078715	2079837
ycfN	putative beta-glucosidase	2038888	2039716
ychM	putative sulphate transporter	3183605	3185285
ycjG	L-Ala-D/L-Glu epimerase	2263455	2264484
ycjI	putative carboxypeptidase	2264480	2265188
ycjJ	putative amino acid-amine transport protein	1960680	1962126
ydaN	putative Zn transport protein	2351030	2352014

ydbK	putative oxidoreductase Fe-S subunit	2357480	2361008
yddG	putative transmembrane protein	2797625	2798504
yddO	putative ABC transport system ATP-binding component	1461324	1462029
yddP	putative ABC transport system oligopeptide translocator	1462015	1462882
yddQ	putative ABC transport system permease component	1462844	1463678
yddR	putative ABC transport system permease component	1463677	1464730
ydeA	putative integral membrane transport protein	2129991	2131251
ydeV	putative kinase	4586918	4588517
ydeW	putative SorC-family transcriptional regulator	4585909	4586881
ydeY	putative ABC transport system permease component	4583172	4584207
ydeZ	putative transport system permease component	4582173	4583172
ydfI	putative mannitol dehydrogenase	2503655	2505119
ydgB	short chain dehydrogenase	2435821	2436544
ydhP	putative transport protein	2922597	2923764
ydiJ	putative oxidase	3091210	3094267
ydiM	putative MFS-family transport protein	871347	872562
yeaE	putative aldehyde reductase	2138734	2139619
yeaS	putative membrane transport protein	2116455	2117094
yeaT	putative LysR-family transcriptional regulator	3607977	3608877
yeaV	putative transport protein	1950117	1951560
yecC	putative amino-acid ABC transport system ATP-binding component	3361776	3362529
yedE	paral putative membrane component of transport system	2417003	2418269
yeeF	putative amino acid transporter protein	3496770	3498129
yeeO	putative MATE family transport protein	3394130	3395600
yegD	putative heat shock protein	3553844	3555197
yegH	putative transmembrane protein	3547260	3548913
yegM	multidrug efflux system membrane protein	3561012	3562251
yegN	multidrug transport protein	3562250	3565373
yegO	multidrug efflux system outer membrane protein	3565373	3568451
yegQ	putative protease	3572380	3573742
yegS	putative diacylglycerol kinase	3573987	3574881
yehU	putative two-compoent regulatory system sensor kinase	3614452	3616141
yehX	putative ABC-type proline/glycine betaine transport system ATP-binding component	3617179	3618127
yehY	putative ABC-type proline/glycine betaine transport system permease component	3618119	3619277
yehZ	putative ABC-type proline/glycine betaine transport system periplasmic component	3619284	3620202
yeiE	putative LysR-family transcriptional regulator	3652691	3653558
yeiG	putative esterase	3647667	3648501
yeiQ	putative mannitol dehydrogenase	3663621	3665097
yfaO	putative NUDIX hydrolase	3729407	3729833

yfaW	putative dehydratase	3723982	3725269
yfaY	competence damage-inducible protein A	3726188	3727385
yfbB	putative enzyme	3733098	3733857
yfcD	putative NUDIX hydrolase	3765541	3766093
yfcG	putative S-transferase	3767583	3768213
yfcH	NAD dependent epimerase/dehydratase family protein	3768659	3769553
yfcX	3-hydroxyacyl-CoA dehydrogenase	3796442	3798587
yfeC	putative negative regulator	3824976	3825330
yfgF	putative cytochrome C-type biogenesis protein	3904997	3907226
yfgK	GTP-binding protein EngA	3919595	3921074
yfhQ	hypothetical tRNA/rRNA methyltransferase	3946635	3947370
ygaA	anaerobic nitric oxide reductase transcription regulator	4133613	4135164
ygaE	putative transcriptional regulator	1071585	1072278
ygaU	putative peptidoglycan-binding protein	2642926	2643376
ygaZ	putative amino acid transport protein	284912	285653
ygbD	nitric oxide reductase	4136796	4137930
ygbO	tRNA pseudouridine synthase D	4198277	4199327
ygeD	putative MFS family efflux protein	4331434	4332628
ygeR	putative lipoprotein	4407420	4408161
ygfF	putative oxidoreductase	4425169	4425913
ygfZ	putative enzyme	4416853	4417837
yggB	component of the MscS mechanosensitive channel	4441712	4442570
yggS	putative enzyme with PLP-binding domain	4463326	4464028
yggW	coproporphyrinogen III oxidase	4465500	4466640
ygiC	putative glutathione-like synthetase	4524274	4525435
ygiD	putative enzyme with dioxygenase domain	4525479	4526271
ygiX	putative 2-component transcriptional regulator	4514197	4514857
ygjG	probable ornithine aminotransferase	4578525	4579932
ygjO	putative enzyme	4590830	4591961
yhaJ	putative LysR-family transcriptional regulator	4609337	4610234
yhbH	probable sigma-54 modulation protein	4686095	4686383
yhbV	putative protease	4642519	4643413
yhbZ	putative GTP-binding factor	4670598	4671777
yhcP	p-hydroxybenzoic acid efflux subunit	4723365	4725333
yheO	putative regulator	4792083	4792818
yhfK	putative dienelactone hydrolase	4803427	4805506
yhfM	putative amino acid/amine transport protein	1131681	1133061
yhgG	putative transcriptional regulator	4844327	4844567
yhhJ	antibiotic transport system permease component	686862	688071
yhhK	putative acyltransferase	4892032	4892416
yhhT	putative PerM-family permease	4910444	4911494

yhhX	putative NAD(P)-binding dehydrogenase	4875800	4876838
yhiH	putative ABC-type multidrug transport system ATPase and permease component	687986	690731
yhjC	putative LysR-family transcriptional regulator	2333692	2334598
yhjJ	putative peptidase	4952955	4954458
yhjL	putative cellulose synthase	4958369	4961849
yhjN	putative cellulose synthase	4962943	4965367
yhjW	putative transmembrane protein	4994108	4995782
yibF	putative S-transferase	5036720	5037329
yicC	putative alpha helix protein	5085347	5086211
yicE	putative purine/xanthine transport protein	5110385	5111792
yicF	DNA ligase	5090405	5092103
yicI	putative alpha-xylosidase	5117913	5120232
yicJ	putative permease	5120244	5121639
yicM	putative inner membrane transport protein	5170985	5172179
yicO	putataive xanthine/uracil/vitamin C permease	3606644	3607988
yidC	inner membrane protein translocase component	5215505	5217152
yidZ	putative LysR-family transcriptional regulator	5220430	5221390
yieC	carbohydrate-specific outer membrane porin in cryptic operon	5227150	5228782
yifK	putative amino acid/amine transport protein	162318	163704
yiiD	putative acetyltransferase	40621	41611
yijO	putative AraC-type regulatory protein	102965	103829
yiuA	putative periplasmic substrate-binding transport protein	1384401	1385520
yjbC	pseudouridine synthase	276977	277853
yjbN	tRNA-dihydrouridine synthase A	304182	305178
yjcS	putative hydrolase	3599882	3600818
yjdJ	putative acyltransferase domain	498295	498826
yjdL	putative POT family di-/tripeptide transport protein	1358532	1360008
yjeF	putative kinase	439889	441416
yjeS	putative electron transport protein	438751	439891
yjfF	inner membrane ABC transport system permease component	494641	495640
yjgB	putative alcohol dehydrogenase	562090	563110
yjgM	putative acetyltransferase	533309	533825
yjgR	putative transport system ATP-binding component	541480	542983
yjiA	putative synthesis protein	715220	716177
yjiY	carbon starvation protein	716501	718652
yjjP	putative membrane protein	755759	756536
yjjV	putative hydrolase	772359	773154
ykgK	putative regulator	1113925	1114468
yliA	putative ABC transport system ATP-binding component	1763304	1765164
yliD	putative ABC transport system permease component	1767706	1768618
yliI	putative dehydrogenase	1772698	1773808

yncB	putative NAD(P)-binding dehydrogenase	2841572	2842610
yncD	probable tonB-dependent receptor	2835885	2837991
yncG	hypothetical GST-like protein	2822465	2823083
yneA	putative LacI-type transcriptional regulator	4581170	4582232
ynfE	putative dimethyl sulfoxide reductase major subunit	2496872	2499308
ynfJ	putative voltage-gated ClC-type chloride channel	2493306	2494602
ynfL	putative LysR-family transcriptional regulator	2490284	2491187
ynfM	putative transport protein	2488903	2490154
ynjC	putative ABC transport system permease component	2159745	2161281
ynjD	putative ABC transport system ATP-binding component	2159209	2159746
ynjE	putative thiosulfate sulfur transferase	2157712	2159020
yoaE	putative transmembrane protein	3277995	3279555
yohG	putative channel/filament proteins	3628041	3629478
yohI	tRNA-dihydrouridine synthase	3635872	3636805
yojL	putative thiamine biosynthesis lipoprotein	3697611	3698664
ypdB	putative 2-component transcriptional regulator	3812710	3813448
ypfI	putative acyl-CoA N-acyltransferase	3882578	3884582
yphG	putative transferase	3955842	3959136
yqeF	acetyl-CoA acetyltransferase	4360585	4361764
yqfA	putative oxidoreductase	4417935	4418595
yqiA	putative hydrolase	4519285	4519861
yqjH	putative FAD-linked/NADP-linked ferredoxin reductase-like protein	4576727	4577504
yraM	putative enzyme	4633355	4635464
yrdD	putative DNA topoisomerase	4762870	4763413
ysgA	putative dienelactone hydrolase	192247	193207
ytfB	putative cell envelope opacity-associated protein A	469864	470560
ytfE	putative cell morphogenesis protein	473228	473891
ytfF	putative cationic amino acid transport protein	473973	474963
ytfG	putative oxidoreductase	475922	476771
ytfL	putative hemolysin-related membrane protein	481504	482842
ytfM	putative outer membrane protein	483869	485603
ytfR	putative D-ribose transport system ATP-binding component	492116	493619
yvfL	putative maltodextrin transport permease	1255675	1256983
znuC	high-affinity zinc uptake system ATP-binding protein	3322527	3323280
zraR	DNA-binding response regulator in two-component regulatory system with ZraS	252669	254001
zraS	sensor kinase of Zn/Pb responsive two-component regulatory system	251288	252683
zwf	glucose-6-phosphate 1-dehydrogenase	3314768	3316244

REFERENCES

Abejew AA, Tamir AS and Kerie MW. (2014) Retrospective analysis of mortalities in a tertiary care hospital in Northeast Ethiopia. *BMC Res Notes* 7: 46.

Abraham EP and Chain E. (1988) An enzyme from bacteria able to destroy penicillin. 1940. *Rev Infect Dis* 10: 677-678.

Adeyemo AA, Gbadegesin RA, Onyemenem TN, et al. (1994) Urinary tract pathogens and antimicrobial sensitivity patterns in children in Ibadan, Nigeria. *Ann Trop Paediatr* 14: 271-274.

Al-Emran HM, Eibach D, Krumkamp R, et al. (2016) A Multicountry Molecular Analysis of Salmonella enterica Serovar Typhi With Reduced Susceptibility to Ciprofloxacin in Sub-Saharan Africa. *Clin Infect Dis* 62 Suppl 1: S42-46.

Alekshun MN and Levy SB. (2007) Molecular mechanisms of antibacterial multidrug resistance. *Cell* 128: 1037-1050.

Altschul SF, Gish W, Miller W, et al. (1990) Basic local alignment search tool. *J Mol Biol* 215: 403-410.

Aminov R. (2017) History of antimicrobial drug discovery: Major classes and health impact. *Biochem Pharmacol* 133: 4-19.

Andersson DI and Hughes D. (2010) Antibiotic resistance and its cost: is it possible to reverse resistance? *Nat Rev Microbiol* 8: 260-271.

Ayoola OO, Adeyemo AA and Osinusi K. (2003) Aetiological agents, clinical features and outcome of septicaemia in infants in Ibadan. *West Afr J Med* 22: 30-34.

Ayoola OO, Adeyemo AA and Osinusi K. (2005) Concurrent bacteraemia and malaria in febrile Nigerian infants. *Trop Doct* 35: 34-36.

Bar-Zeev N, Mtunthama N, Gordon SB, et al. (2015) Minimum incidence of adult invasive pneumococcal disease in Blantyre, Malawi an urban african setting: a hospital based prospective cohort study. *PLoS One* 10: e0128738.

Barrow GI, Feltham, R.K.A. (1993) Cowan and Steel's Manual for the Identification of Medical Bacteria. Third ed.: Cambridge University Press.

Bialek-Davenet S, Criscuolo A, Ailloud F, et al. (2014) Genomic definition of hypervirulent and multidrug-resistant Klebsiella pneumoniae clonal groups. *Emerg Infect Dis* 20: 1812-1820.

Blomberg B, Jureen R, Manji KP, et al. (2005) High rate of fatal cases of pediatric septicemia caused by gram-negative bacteria with extended-spectrum beta-lactamases in Dar es Salaam, Tanzania. *J Clin Microbiol* 43: 745-749.

Boetzer M, Henkel CV, Jansen HJ, et al. (2011) Scaffolding pre-assembled contigs using SSPACE. *Bioinformatics* 27: 578-579.

Bonnet R. (2004) Growing group of extended-spectrum beta-lactamases: the CTX-M enzymes. *Antimicrob Agents Chemother* 48: 1-14.

Bouza E and Cercenado E. (2002) Klebsiella and enterobacter: antibiotic resistance and treatment implications. *Semin Respir Infect* 17: 215-230.

Bowers JR, Kitchel B, Driebe EM, et al. (2015) Genomic Analysis of the Emergence and Rapid Global Dissemination of the Clonal Group 258 Klebsiella pneumoniae Pandemic. *PLoS One* 10: e0133727.

Breurec S, Guessennd N, Timinouni M, et al. (2013) Klebsiella pneumoniae resistant to third-generation cephalosporins in five African and two Vietnamese major towns: multiclonal population structure with two major international clonal groups, CG15 and CG258. *Clin Microbiol Infect* 19: 349-355.

Brink AJ, Botha RF, Poswa X, et al. (2012a) Antimicrobial susceptibility of gram-negative pathogens isolated from patients with complicated intra-abdominal infections in South African hospitals (SMART Study 2004-2009): impact of the new carbapenem breakpoints. *Surg Infect (Larchmt)* 13: 43-49.

Brink AJ, Coetzee J, Clay CG, et al. (2012b) Emergence of New Delhi metallo-beta-lactamase (NDM-1) and Klebsiella pneumoniae carbapenemase (KPC-2) in South Africa. *J Clin Microbiol* 50: 525-527.

Brisse S and Verhoef J. (2001) Phylogenetic diversity of Klebsiella pneumoniae and Klebsiella oxytoca clinical isolates revealed by randomly amplified polymorphic DNA, gyrA and parC genes sequencing and automated ribotyping. *Int J Syst Evol Microbiol* 51: 915-924.

Brown L, Wolf JM, Prados-Rosales R, et al. (2015) Through the wall: extracellular vesicles in Gram-positive bacteria, mycobacteria and fungi. *Nat Rev Microbiol* 13: 620-630.

Caboche S, Audebert C, Lemoine Y, et al. (2014) Comparison of mapping algorithms used in high-throughput sequencing: application to Ion Torrent data. *BMC Genomics* 15: 264.

Calbo E and Garau J. (2015) The changing epidemiology of hospital outbreaks due to ESBL-producing Klebsiella pneumoniae: the CTX-M-15 type consolidation. *Future Microbiol* 10: 1063-1075.

Campbell EA, Korzheva N, Mustaev A, et al. (2001) Structural mechanism for rifampicin inhibition of bacterial rna polymerase. *Cell* 104: 901-912.

Carattoli A, Zankari E, Garcia-Fernandez A, et al. (2014) In silico detection and typing of plasmids using PlasmidFinder and plasmid multilocus sequence typing. *Antimicrob Agents Chemother* 58: 3895-3903.

Castanheira M, Farrell SE, Krause KM, et al. (2014) Contemporary diversity of beta-lactamases among Enterobacteriaceae in the nine U.S. census regions and ceftazidime-avibactam activity tested against isolates producing the most prevalent beta-lactamase groups. *Antimicrob Agents Chemother* 58: 833-838.

Central Intelligence Agancy. (2017) Malawi. *The World Factbook*. (accessed 2017).

Chang-Ro Lee JHL, Kwang Seung Park, Young Bae Kim, Byeong Chul Jeong and Sang Hee Lee. (2016) Global Dissemination of Carbapenemase-Producing Klebsiella pneumoniae: Epidemiology, Genetic Context, Treatment Options, and Detection Methods. *Frontiers in Microbiology* 7.

Chapman SJ and Hill AV. (2012) Human genetic susceptibility to infectious disease. *Nat Rev Genet* 13: 175-188.

Chattaway MA, Aboderin AO, Fashae K, et al. (2016) Fluoroquinolone-Resistant Enteric Bacteria in Sub-Saharan Africa: Clones, Implications and Research Needs. *Front Microbiol* 7: 558.

Chaudhuri RR and Henderson IR. (2012) The evolution of the Escherichia coli phylogeny. *Infect Genet Evol* 12: 214-226.

Chen LF, Freeman JT, Nicholson B, et al. (2014) Widespread dissemination of CTX-M-15 genotype extended-spectrum-beta-lactamase-producing enterobacteriaceae among patients presenting to community hospitals in the southeastern United States. *Antimicrob Agents Chemother* 58: 1200-1202.

Chimalizeni Y, Kawaza K and Molyneux E. (2010) The epidemiology and management of non typhoidal salmonella infections. *Adv Exp Med Biol* 659: 33-46.

Cho H, Uehara T and Bernhardt TG. (2014) Beta-lactam antibiotics induce a lethal malfunctioning of the bacterial cell wall synthesis machinery. *Cell* 159: 1300-1311.

Chopra I and Roberts M. (2001) Tetracycline antibiotics: mode of action, applications, molecular biology, and epidemiology of bacterial resistance. *Microbiol Mol Biol Rev* 65: 232-260 ; second page, table of contents.

Chopra S and Galande A. (2011) A fluoroquinolone-resistant Acinetobacter baumannii without the quinolone resistance-determining region mutations. *J Antimicrob Chemother* 66: 2668-2670.

Chung The H, Karkey A, Pham Thanh D, et al. (2015) A high-resolution genomic analysis of multidrug-resistant hospital outbreaks of Klebsiella pneumoniae. *EMBO Mol Med* 7: 227-239.

Clermont O, Christenson JK, Denamur E, et al. (2013) The Clermont Escherichia coli phylo-typing method revisited: improvement of specificity and detection of new phylo-groups. *Environ Microbiol Rep* 5: 58-65.

Coelho FC, Codeco CT and Gomes MG. (2011) A Bayesian framework for parameter estimation in dynamical models. *PLoS One* 6: e19616.

Corander J and Tang J. (2007) Bayesian analysis of population structure based on linked molecular information. *Math Biosci* 205: 19-31.

Cornick JE and Bentley SD. (2012) Streptococcus pneumoniae: the evolution of antimicrobial resistance to beta-lactams, fluoroquinolones and macrolides. *Microbes Infect* 14: 573-583.

Croucher NJ, Page AJ, Connor TR, et al. (2015) Rapid phylogenetic analysis of large samples of recombinant bacterial whole genome sequences using Gubbins. *Nucleic Acids Res* 43: e15.

Croxen MA, Law RJ, Scholz R, et al. (2013) Recent advances in understanding enteric pathogenic Escherichia coli. *Clin Microbiol Rev* 26: 822-880.

D'Andrea MM, Arena F, Pallecchi L, et al. (2013) CTX-M-type beta-lactamases: a successful story of antibiotic resistance. *Int J Med Microbiol* 303: 305-317.

Davies J and Davies D. (2010) Origins and evolution of antibiotic resistance. *Microbiol Mol Biol Rev* 74: 417-433.

Daza P, Banda R, Misoya K, et al. (2006) The impact of routine infant immunization with Haemophilus influenzae type b conjugate vaccine in Malawi, a country with high human immunodeficiency virus prevalence. *Vaccine* 24: 6232-6239.

Demain AL and Sanchez S. (2009) Microbial drug discovery: 80 years of progress. *J Antibiot (Tokyo)* 62: 5-16.

Denning G, Kabambe P, Sanchez P, et al. (2009) Input subsidies to improve smallholder maize productivity in Malawi: toward an african green revolution. *PLoS Biol* 7: e23.

Desta K, Woldeamanuel Y, Azazh A, et al. (2016) High Gastrointestinal Colonization Rate with Extended-Spectrum beta-Lactamase-Producing Enterobacteriaceae in Hospitalized Patients: Emergence of Carbapenemase-Producing K. pneumoniae in Ethiopia. *PLoS One* 11: e0161685.

Didelot X and Maiden MC. (2010) Impact of recombination on bacterial evolution. *Trends Microbiol* 18: 315-322.

Dramowski A, Cotton MF, Rabie H, et al. (2015) Trends in paediatric bloodstream infections at a South African referral hospital. *BMC Pediatr* 15: 33.

Drummond AJ and Rambaut A. (2007) BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evol Biol* 7: 214.

Duchene S, Holt KE, Weill FX, et al. (2016) Genome-scale rates of evolutionary change in bacteria. *Microb Genom* 2: e000094.

Eibach D, Campos CB, Krumkamp R, et al. (2016) Extended spectrum beta-lactamase producing Enterobacteriaceae causing bloodstream infections in rural Ghana, 2007-2012. *Int J Med Microbiol* 306: 249-254.

Everett DB, Mukaka M, Denis B, et al. (2011) Ten years of surveillance for invasive Streptococcus pneumoniae during the era of antiretroviral scale-up and cotrimoxazole prophylaxis in Malawi. *PLoS One* 6: e17765.

Feasey N. (2014) An investigation into clinical, epidemiological and genomic changes in epidemic Salmonella bloodstream infection in Malawi. University of Liverpool.

Feasey NA, Cain AK, Msefula CL, et al. (2014a) Drug resistance in Salmonella enterica ser. Typhimurium bloodstream infection, Malawi. *Emerg Infect Dis* 20: 1957-1959.

Feasey NA, Everett D, Faragher EB, et al. (2015a) Modelling the Contributions of Malaria, HIV, Malnutrition and Rainfall to the Decline in Paediatric Invasive Non-typhoidal Salmonella Disease in Malawi. *PLoS Negl Trop Dis* 9: e0003979.

Feasey NA, Gaskell K, Wong V, et al. (2015b) Rapid emergence of multidrug resistant, H58-lineage Salmonella typhi in Blantyre, Malawi. *PLoS Negl Trop Dis* 9: e0003748.

Feasey NA, Houston A, Mukaka M, et al. (2014b) A reduction in adult blood stream infection and case fatality at a large African hospital following antiretroviral therapy roll-out. *PLoS One* 9: e92226.

Feasey NA, Masesa C, Jassi C, et al. (2015c) Three Epidemics of Invasive Multidrug-Resistant Salmonella Bloodstream Infection in Blantyre, Malawi, 1998-2014. *Clin Infect Dis* 61 Suppl 4: S363-371.

Ferjani S, Saidani M, Ennigrou S, et al. (2014) Multidrug resistance and high virulence genotype in uropathogenic Escherichia coli due to diffusion of ST131 clonal group producing CTX-M-15: an emerging problem in a Tunisian hospital. *Folia Microbiol (Praha)* 59: 257-262.

Fortuin-de Smidt MC, Singh-Moodley A, Badat R, et al. (2015) Staphylococcus aureus bacteraemia in Gauteng academic hospitals, South Africa. *Int J Infect Dis* 30: 41-48.

Gangoue-Pieboji J, Bedenic B, Koulla-Shiro S, et al. (2005a) Extended-spectrum-beta-lactamase-producing Enterobacteriaceae in Yaounde, Cameroon. *J Clin Microbiol* 43: 3273-3277.

Gangoue-Pieboji J, Koulla-Shiro S, Ngassam P, et al. (2006) Antimicrobial activity against gram negative bacilli from Yaounde Central Hospital, Cameroon. *Afr Health Sci* 6: 232-235.

Gangoue-Pieboji J, Miriagou V, Vourli S, et al. (2005b) Emergence of CTX-M-15-producing enterobacteria in Cameroon and characterization of a blaCTX-M-15-carrying element. *Antimicrob Agents Chemother* 49: 441-443.

Garza-Ramos U, Moreno-Dominguez S, Hernandez-Castro R, et al. (2016) Identification and Characterization of Imipenem-Resistant Klebsiella pneumoniae and Susceptible Klebsiella variicola Isolates Obtained from the Same Patient. *Microb Drug Resist* 22: 179-184.

Gordon MA, Graham SM, Walsh AL, et al. (2008) Epidemics of invasive Salmonella enterica serovar enteritidis and S. enterica Serovar typhimurium infection associated with multidrug resistance among adults and children in Malawi. *Clin Infect Dis* 46: 963-969.

Gordon MA, Walsh AL, Chaponda M, et al. (2001) Bacteraemia and mortality among adult medical admissions in Malawi--predominance of non-typhi salmonellae and Streptococcus pneumoniae. *J Infect* 42: 44-49.

Gordon SB, Walsh AL, Chaponda M, et al. (2000) Bacterial meningitis in Malawian adults: pneumococcal disease is common, severe, and seasonal. *Clin Infect Dis* 31: 53-57.

Govinden U, Mocktar C, Moodley P, et al. (2006) CTX-M-37 in Salmonella enterica serotype Isangi from Durban, South Africa. *Int J Antimicrob Agents* 28: 288-291.

Gray KJ, Wilson, L.K., A. Phiri, A., Corkill, J.E., French, N., et al., (2006) (2006) Identification and characterization of ceftriaxone resistance and extended-spectrum beta-lactamases in Malawian bacteraemic Enterobacteriaceae. *J Antimicrob Chemother* 57: 661-665.

Grohn YT, Carson C, Lanzas C, et al. (2017) A proposed analytic framework for determining the impact of an antimicrobial resistance intervention. *Anim Health Res Rev*: 1-25.

Hakenbeck R, Bruckner R, Denapaite D, et al. (2012) Molecular mechanisms of beta-lactam resistance in Streptococcus pneumoniae. *Future Microbiol* 7: 395-410.

Hall BG and Barlow M. (2005) Revised Ambler classification of {beta}-lactamases. *J Antimicrob Chemother* 55: 1050-1051.

Harbarth S and Samore MH. (2005) Antimicrobial resistance determinants and future control. *Emerg Infect Dis* 11: 794-801.

Hawkey PM. (2003) Mechanisms of quinolone action and microbial response. *J Antimicrob Chemother* 51 Suppl 1: 29-35.

Hendriksen RS, Leekitcharoenphon P, Lukjancenko O, et al. (2015) Genomic signature of multidrug-resistant Salmonella enterica serovar typhi isolates

related to a massive outbreak in Zambia between 2010 and 2012. *J Clin Microbiol* 53: 262-272.
Holmes AH, Moore LS, Sundsfjord A, et al. (2016) Understanding the mechanisms and drivers of antimicrobial resistance. *Lancet* 387: 176-187.
Holmes CB, Losina E, Walensky RP, et al. (2003) Review of human immunodeficiency virus type 1-related opportunistic infections in sub-Saharan Africa. *Clin Infect Dis* 36: 652-662.
Holt KE, Wertheim H, Zadoks RN, et al. (2015) Genomic analysis of diversity, population structure, virulence, and antimicrobial resistance in Klebsiella pneumoniae, an urgent threat to public health. *Proc Natl Acad Sci U S A* 112: E3574-3581.
Hooper DC and Jacoby GA. (2015a) Mechanisms of drug resistance: quinolone resistance. *Ann N Y Acad Sci*.
Hooper DC and Jacoby GA. (2015b) Mechanisms of drug resistance: quinolone resistance. *Ann N Y Acad Sci* 1354: 12-31.
Hu L, Zhong Q, Tu J, et al. (2013) Emergence of blaNDM-1 among Klebsiella pneumoniae ST15 and novel ST1031 clinical isolates in China. *Diagn Microbiol Infect Dis* 75: 373-376.
Huson DH. (1998) SplitsTree: analyzing and visualizing evolutionary data. *Bioinformatics* 14: 68-73.
Huson DH and Scornavacca C. (2012) Dendroscope 3: an interactive tool for rooted phylogenetic trees and networks. *Syst Biol* 61: 1061-1067.
Huson MA, Stolp SM, van der Poll T, et al. (2014) Community-acquired bacterial bloodstream infections in HIV-infected patients: a systematic review. *Clin Infect Dis* 58: 79-92.
Hussain A, Ranjan A, Nandanwar N, et al. (2014) Genotypic and phenotypic profiles of Escherichia coli isolates belonging to clinical sequence type 131 (ST131), clinical non-ST131, and fecal non-ST131 lineages from India. *Antimicrob Agents Chemother* 58: 7240-7249.
Infectious Diseases Society of America, Spellberg B, Blaser M, et al. (2011) Combating antimicrobial resistance: policy recommendations to save lives. *Clin Infect Dis* 52 Suppl 5: S397-428.
Interagency Task Force on Antimicrobial Resistance. (2012) A public Health Action Plan to Combat Antimicrobial Resistance: 2012 Update. (accessed 2017).
Jackson RW, Vinatzer B, Arnold DL, et al. (2011) The influence of the accessory genome on bacterial pathogen evolution. *Mob Genet Elements* 1: 55-65.
Jacobson RK, Manesen MR, Moodley C, et al. (2015) Molecular characterisation and epidemiological investigation of an outbreak of blaOXA-181 carbapenemase-producing isolates of Klebsiella pneumoniae in South Africa. *S Afr Med J* 105: 1030-1035.
Jacoby GA, Strahilevitz J and Hooper DC. (2014) Plasmid-mediated quinolone resistance. *Microbiol Spectr* 2.
Johnson JR and Russo TA. (2002) Extraintestinal pathogenic Escherichia coli: "the other bad E coli". *J Lab Clin Med* 139: 155-162.
Jones RN. (2010) Microbial etiologies of hospital-acquired bacterial pneumonia and ventilator-associated bacterial pneumonia. *Clin Infect Dis* 51 Suppl 1: S81-87.

Jonson JR, Russo, T.A. (2005) Molecular epidemiology of extraintesinal pathogenic (uropathogenic) *Escherichia coli. International Journal of Medical Microbiology* 295: 383--404.

Jovetic S, Zhu Y, Marcone GL, et al. (2010) beta-Lactam and glycopeptide antibiotics: first and last line of defense? *Trends Biotechnol* 28: 596-604.

Kaper JB, Nataro JP and Mobley HL. (2004) Pathogenic Escherichia coli. *Nat Rev Microbiol* 2: 123-140.

Kariuki S, Revathi G, Corkill J, et al. (2007) Escherichia coli from community-acquired urinary tract infections resistant to fluoroquinolones and extended-spectrum beta-lactams. *J Infect Dev Ctries* 1: 257-262.

Kariuki S, Revathi G, Kariuki N, et al. (2006) Invasive multidrug-resistant non-typhoidal Salmonella infections in Africa: zoonotic or anthroponotic transmission? *J Med Microbiol* 55: 585-591.

Kariuki S, Revathi G, Kiiru J, et al. (2010) Typhoid in Kenya is associated with a dominant multidrug-resistant Salmonella enterica serovar Typhi haplotype that is also widespread in Southeast Asia. *J Clin Microbiol* 48: 2171-2176.

Kaye KS, Engemann JJ, Fraimow HS, et al. (2004) Pathogens resistant to antimicrobial agents: epidemiology, molecular mechanisms, and clinical management. *Infect Dis Clin North Am* 18: 467-511, viii.

Keddy KH, Smith AM, Sooka A, et al. (2010) Fluoroquinolone-resistant typhoid, South Africa. *Emerg Infect Dis* 16: 879-880.

Keeling MJR, P. (2008) *Modeling Infectious Diseases in Humans and Animals,* Princeton, New Jersey, USA.: Princeton University Press.

King AA, Nguyen, D. Ionides, E.L. . (2016) Statistical Inference for Partially Observed Markov Processes via the R Package pomp. *Journal of Statistical Software* 69.

Kohanski MA, Dwyer DJ and Collins JJ. (2010) How antibiotics kill bacteria: from targets to networks. *Nat Rev Microbiol* 8: 423-435.

Kohler CD and Dobrindt U. (2011) What defines extraintestinal pathogenic Escherichia coli? *Int J Med Microbiol* 301: 642-647.

Kotra LP, Haddad J and Mobashery S. (2000) Aminoglycosides: perspectives on mechanisms of action and resistance and strategies to counter resistance. *Antimicrob Agents Chemother* 44: 3249-3256.

Kruger T, Szabo D, Keddy KH, et al. (2004) Infections with nontyphoidal Salmonella species producing TEM-63 or a novel TEM enzyme, TEM-131, in South Africa. *Antimicrob Agents Chemother* 48: 4263-4270.

Lamikanra A, Crowe JL, Lijek RS, et al. (2011) Rapid evolution of fluoroquinolone-resistant Escherichia coli in Nigeria is temporally associated with fluoroquinolone use. *BMC Infect Dis* 11: 312.

Laupland KB, Gregson, D.B., Church, D.L., Ross, T., Pitout, J.D.D. (2008) Incidence, risk factors and outcomes of Escherichia coli bloodstream infections in a large Canadian region. *Clinical Microbiology and Infection* 14: 1041-1047.

Lautenbach E, Strom BL, Bilker WB, et al. (2001) Epidemiological investigation of fluoroquinolone resistance in infections due to extended-spectrum beta-lactamase-producing Escherichia coli and Klebsiella pneumoniae. *Clin Infect Dis* 33: 1288-1294.

Laxminarayan R, Duse A, Wattal C, et al. (2013) Antibiotic resistance-the need for global solutions. *Lancet Infect Dis* 13: 1057-1098.

Laxminarayan R, Matsoso P, Pant S, et al. (2016) Access to effective antimicrobials: a worldwide challenge. *Lancet* 387: 168-175.

Leopold SJ, van Leth F, Tarekegn H, et al. (2014) Antimicrobial drug resistance among clinically relevant bacterial isolates in sub-Saharan Africa: a systematic review. *J Antimicrob Chemother* 69: 2337-2353.

Letunic I and Bork P. (2016) Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. *Nucleic Acids Res* 44: W242-245.

Li H and Durbin R. (2009) Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25: 1754-1760.

Li XZ, Ma D, Livermore DM, et al. (1994) Role of efflux pump(s) in intrinsic resistance of Pseudomonas aeruginosa: active efflux as a contributing factor to beta-lactam resistance. *Antimicrob Agents Chemother* 38: 1742-1752.

Ligon BL. (2004a) Penicillin: its discovery and early development. *Semin Pediatr Infect Dis* 15: 52-57.

Ligon BL. (2004b) Sir Alexander Fleming: Scottish researcher who discovered penicillin. *Semin Pediatr Infect Dis* 15: 58-64.

Ligon BL. (2004c) Sir Howard Walter Florey--the force behind the development of penicillin. *Semin Pediatr Infect Dis* 15: 109-114.

Liu L, Johnson HL, Cousens S, et al. (2012) Global, regional, and national causes of child mortality: an updated systematic analysis for 2010 with time trends since 2000. *Lancet* 379: 2151-2161.

Lonchel CM, Meex C, Gangoue-Pieboji J, et al. (2012) Proportion of extended-spectrum ss-lactamase-producing Enterobacteriaceae in community setting in Ngaoundere, Cameroon. *BMC Infect Dis* 12: 53.

Lunguya O, Lejon V, Phoba MF, et al. (2012) Salmonella typhi in the democratic republic of the congo: fluoroquinolone decreased susceptibility on the rise. *PLoS Negl Trop Dis* 6: e1921.

Madurga S, Sanchez-Cespedes J, Belda I, et al. (2008) Mechanism of binding of fluoroquinolones to the quinolone resistance-determining region of DNA gyrase: towards an understanding of the molecular basis of quinolone resistance. *Chembiochem* 9: 2081-2086.

Magiorakos AP, Srinivasan A, Carey RB, et al. (2012) Multidrug-resistant, extensively drug-resistant and pandrug-resistant bacteria: an international expert proposal for interim standard definitions for acquired resistance. *Clin Microbiol Infect* 18: 268-281.

Maiden MC. (2006) Multilocus sequence typing of bacteria. *Annu Rev Microbiol* 60: 561-588.

Maiden MC, Bygraves JA, Feil E, et al. (1998) Multilocus sequence typing: a portable approach to the identification of clones within populations of pathogenic microorganisms. *Proc Natl Acad Sci U S A* 95: 3140-3145.

Maina D, Omuse G, Revathi G, et al. (2016) Spectrum of Microbial Diseases and Resistance Patterns at a Private Teaching Hospital in Kenya: Implications for Clinical Practice. *PLoS One* 11: e0147659.

Makoka MH, Miller WC, Hoffman IF, et al. (2012) Bacterial infections in Lilongwe, Malawi: aetiology and antibiotic resistance. *BMC Infect Dis* 12: 67.

Makombe SD, Jahn A, Tweya H, et al. (2007) A national survey of the impact of rapid scale-up of antiretroviral therapy on health-care workers in Malawi: effects on human resources and survival. *Bull World Health Organ* 85: 851-857.

Mandomando I, Sigauque B, Morais L, et al. (2010) Antimicrobial drug resistance trends of bacteremia isolates in a rural hospital in southern Mozambique. *Am J Trop Med Hyg* 83: 152-157.

Marttinen P, Hanage WP, Croucher NJ, et al. (2012) Detection of recombination events in bacterial genomes from large population samples. *Nucleic Acids Res* 40: e6.

Masters PA, O'Bryan TA, Zurlo J, et al. (2003) Trimethoprim-sulfamethoxazole revisited. *Arch Intern Med* 163: 402-410.

Mathanga DP, Walker ED, Wilson ML, et al. (2012) Malaria control in Malawi: current status and directions for the future. *Acta Trop* 121: 212-217.

Matsumura Y, Johnson JR, Yamamoto M, et al. (2015) CTX-M-27- and CTX-M-14-producing, ciprofloxacin-resistant Escherichia coli of the H30 subclonal group within ST131 drive a Japanese regional ESBL epidemic. *J Antimicrob Chemother* 70: 1639-1649.

Mauchaza K, Madzimbamuto FD and Waner S. (2016) Methicillin-resistant Staphylococcus aureus in Zimbabwe. *Ghana Med J* 50: 68-71.

Mazel D. (2006) Integrons: agents of bacterial evolution. *Nat Rev Microbiol* 4: 608-620.

McKenna A, Hanna M, Banks E, et al. (2010) The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res* 20: 1297-1303.

Miller CP and Bohnhoff M. (1950) The development of bacterial resistance to chemotherapeutic agents. *Annu Rev Microbiol* 4: 210-222.

Mingeot-Leclercq MP, Glupczynski Y and Tulkens PM. (1999) Aminoglycosides: activity and resistance. *Antimicrob Agents Chemother* 43: 727-737.

Moges AF, Genetu A and Mengistu G. (2002) Antibiotic sensitivities of common bacterial pathogens in urinary tract infections at Gondar Hospital, Ethiopia. *East Afr Med J* 79: 140-142.

MoH. (2009) Malawi Standard Treatment Guidelines. Fourth ed.

Molyneux EM, Mankhambo LA, Phiri A, et al. (2009) The outcome of non-typhoidal salmonella meningitis in Malawian children, 1997-2006. *Ann Trop Paediatr* 29: 13-22.

Moradigaravand D, Martin V, Peacock SJ, et al. (2017) Evolution and Epidemiology of Multidrug-Resistant Klebsiella pneumoniae in the United Kingdom and Ireland. *MBio* 8.

Morpeth SC, Ramadhani HO and Crump JA. (2009) Invasive non-Typhi Salmonella disease in Africa. *Clin Infect Dis* 49: 606-611.

Mshana SE, Falgenhauer L, Mirambo MM, et al. (2016) Predictors of blaCTX-M-15 in varieties of Escherichia coli genotypes from humans in community settings in Mwanza, Tanzania. *BMC Infect Dis* 16: 187.

Mshana SE, Hain T, Domann E, et al. (2013a) Predominance of Klebsiella pneumoniae ST14 carrying CTX-M-15 causing neonatal sepsis in Tanzania. *BMC Infect Dis* 13: 466.

Mshana SE, Matee M and Rweyemamu M. (2013b) Antimicrobial resistance in human and animal pathogens in Zambia, Democratic Republic of Congo, Mozambique and Tanzania: an urgent need of a sustainable surveillance system. *Ann Clin Microbiol Antimicrob* 12: 28.

Munita JM and Arias CA. (2016) Mechanisms of Antibiotic Resistance. *Microbiol Spectr* 4.

Mwagomba B, Zachariah R, Massaquoi M, et al. (2010) Mortality reduction associated with HIV/AIDS care and antiretroviral treatment in rural Malawi: evidence from registers, coffin sales and funerals. *PLoS One* 5: e10452.

Nadalin F, Vezzi F and Policriti A. (2012) GapFiller: a de novo assembly approach to fill the gap within paired reads. *BMC Bioinformatics* 13 Suppl 14: S8.

Nagelkerke N, de Vlas SJ, Jha P, et al. (2009) Heterogeneity in host HIV susceptibility as a potential contributor to recent HIV prevalence declines in Africa. *AIDS* 23: 125-130.

Nakamura T, Komatsu M, Yamasaki K, et al. (2012) Epidemiology of Escherichia coli, Klebsiella species, and Proteus mirabilis strains producing extended-spectrum beta-lactamases from clinical samples in the Kinki Region of Japan. *Am J Clin Pathol* 137: 620-626.

Nakano S, Fujisawa T, Ito Y, et al. (2016) Serotypes, antimicrobial susceptibility, and molecular epidemiology of invasive and non-invasive Streptococcus pneumoniae isolates in paediatric patients after the introduction of 13-valent conjugate vaccine in a nationwide surveillance study conducted in Japan in 2012-2014. *Vaccine* 34: 67-76.

National AIDS Commission. (2015) Malawi AIDS Response Progress Report 2015.

National statistical Office (NSO). Population Projections for Malawi.

National Statistical Office (NSO) and ICF. (2017) Malawi Demographic and Health Survey 2015-2016. Zomba, Malawi and Rockville, Maryland, USA: NSO and ICF.

National Statistical Office (NSO) and ICF Macro. (2011) Malawi Demographic and Health Survey 2010. In: Macro NaI (ed). Zomba, Malawi and Maryland, USA.

National Statistical Office (NSO) and ORC Macro. (2005) Malawi Demographic and Health Survey 2004. Calverton, Maryland: NSO and ORC Macro.

Nelson C. (2012) Controlling the typhoid epidemic plaguing sub-Saharan Africa. *The Atlantic*.

Nicolas-Chanoine MH, Bertrand X and Madec JY. (2014) Escherichia coli ST131, an intriguing clonal group. *Clin Microbiol Rev* 27: 543-574.

Nicolas-Chanoine MH, Blanco J, Leflon-Guibout V, et al. (2008) Intercontinental emergence of Escherichia coli clone O25:H4-ST131 producing CTX-M-15. *J Antimicrob Chemother* 61: 273-281.

Nikaido H. (2009) Multidrug resistance in bacteria. *Annu Rev Biochem* 78: 119-146.

Normark BH and Normark S. (2002) Evolution and spread of antibiotic resistance. *J Intern Med* 252: 91-106.

Nyasulu JY, Nyasulu, P. (2011) BARRIERS TO THE UPTAKE OF PREVENTION OF MOTHER TO CHILD TRANSMISSION
(PMTCT) SERVICES IN RURAL BLANTYRE AND BALAKA DISTRICTS, MALAWI *Journal of Rural and Tropical Public Health* 10: 48-52.

O'Brien TF. (2002) Emergence, spread, and environmental effect of antimicrobial resistance: how use of an antimicrobial anywhere can increase resistance to any antimicrobial anywhere else. *Clin Infect Dis* 34 Suppl 3: S78-84.

O'Neill TRoARCbJ. (2014) Antimicrobial Resistance Tackling a crisis for the health and wealth of nations

Obeng-Nkrumah N, Labi AK, Addison NO, et al. (2016) Trends in paediatric and adult bloodstream infections at a Ghanaian referral hospital: a retrospective study. *Ann Clin Microbiol Antimicrob* 15: 49.

Obi CL, Tarupiwa A and Simango C. (1996) Scope of urinary pathogens isolated in the Public Health Bacteriology Laboratory, Harare: antibiotic susceptibility patterns of isolates and incidence of haemolytic bacteria. *Cent Afr J Med* 42: 244-249.

Okoro CK, Kingsley RA, Connor TR, et al. (2012) Intracontinental spread of human invasive Salmonella Typhimurium pathovariants in sub-Saharan Africa. *Nat Genet* 44: 1215-1221.

Oteo J, Perez-Vazquez M, Bautista V, et al. (2016) The spread of KPC-producing Enterobacteriaceae in Spain: WGS analysis of the emerging high-risk clones of Klebsiella pneumoniae ST11/KPC-2, ST101/KPC-2 and ST512/KPC-3. *J Antimicrob Chemother*.

Ouedraogo AS, Sanou M, Kissou A, et al. (2016) High prevalence of extended-spectrum β-lactamase producing enterobacteriaceae among clinical isolates in Burkina Faso. *BMC Infect Dis* 16: 326.

Page AJ, Cummins CA, Hunt M, et al. (2015) Roary: rapid large-scale prokaryote pan genome analysis. *Bioinformatics* 31: 3691-3693.

Page AJ, Taylor B, Delaney AJ, et al. (2016) SNP-sites: rapid efficient extraction of SNPs from multi-FASTA alignments. *Microbial Genomics*.

Paterson DL. (2006) Resistance in gram-negative bacteria: enterobacteriaceae. *Am J Med* 119: S20-28; discussion S62-70.

Peirano G, Richardson D, Nigrin J, et al. (2010) High prevalence of ST131 isolates producing CTX-M-15 and CTX-M-14 among extended-spectrum-beta-lactamase-producing Escherichia coli isolates from Canada. *Antimicrob Agents Chemother* 54: 1327-1330.

Peirano G, van Greune CH and Pitout JD. (2011) Characteristics of infections caused by extended-spectrum beta-lactamase-producing Escherichia coli from community hospitals in South Africa. *Diagn Microbiol Infect Dis* 69: 449-453.

Peltola H. (2001) Burden of meningitis and other severe bacterial infections of children in africa: implications for prevention. *Clin Infect Dis* 32: 64-75.

Perez-Llarena FJ and Bou G. (2009) Beta-lactamase inhibitors: the story so far. *Curr Med Chem* 16: 3740-3765.

Perovic O, Iyaloo S, Kularatne R, et al. (2015) Prevalence and Trends of Staphylococcus aureus Bacteraemia in Hospitalized Patients in South Africa,

2010 to 2012: Laboratory-Based Surveillance Mapping of Antimicrobial Resistance and Molecular Epidemiology. *PLoS One* 10: e0145429.

Petty NK, Ben Zakour NL, Stanton-Cook M, et al. (2014) Global dissemination of a multidrug resistant Escherichia coli clone. *Proc Natl Acad Sci U S A* 111: 5694-5699.

Philippon A, Labia R and Jacoby G. (1989) Extended-spectrum beta-lactamases. *Antimicrob Agents Chemother* 33: 1131-1136.

Pitout JD. (2012) Extraintestinal Pathogenic Escherichia coli: A Combination of Virulence with Antibiotic Resistance. *Front Microbiol* 3: 9.

Pitout JD, Nordmann P, Laupland KB, et al. (2005) Emergence of Enterobacteriaceae producing extended-spectrum beta-lactamases (ESBLs) in the community. *J Antimicrob Chemother* 56: 52-59.

Pitout JD, Nordmann P and Poirel L. (2015) Carbapenemase-Producing Klebsiella pneumoniae, a Key Pathogen Set for Global Nosocomial Dominance. *Antimicrob Agents Chemother* 59: 5873-5884.

Pitout JD, Thomson KS, Hanson ND, et al. (1998) beta-Lactamases responsible for resistance to expanded-spectrum cephalosporins in Klebsiella pneumoniae, Escherichia coli, and Proteus mirabilis isolates recovered in South Africa. *Antimicrob Agents Chemother* 42: 1350-1354.

Pitzer VE, Feasey NA, Msefula C, et al. (2015) Mathematical Modeling to Assess the Drivers of the Recent Emergence of Typhoid Fever in Blantyre, Malawi. *Clin Infect Dis* 61 Suppl 4: S251-258.

Podschun R and Ullmann U. (1998) Klebsiella spp. as nosocomial pathogens: epidemiology, taxonomy, typing methods, and pathogenicity factors. *Clin Microbiol Rev* 11: 589-603.

Poirel L, Revathi G, Bernabeu S, et al. (2011) Detection of NDM-1-producing Klebsiella pneumoniae in Kenya. *Antimicrob Agents Chemother* 55: 934-936.

Powers JH. (2004) Antimicrobial drug development--the past, the present, and the future. *Clin Microbiol Infect* 10 Suppl 4: 23-31.

Price LB, Johnson JR, Aziz M, et al. (2013) The epidemic of extended-spectrum-beta-lactamase-producing Escherichia coli ST131 is driven by a single highly pathogenic subclone, H30-Rx. *MBio* 4: e00377-00313.

R Core Team. (2014) R: A language and environment for statistical computing. . Vienna, Austria: R Foundation for Statistical Computing.

Raji MA, Jamal W, Ojemeh O, et al. (2015) Sequence analysis of genes mediating extended-spectrum beta-lactamase (ESBL) production in isolates of Enterobacteriaceae in a Lagos Teaching Hospital, Nigeria. *BMC Infect Dis* 15: 259.

Rambaut A. (2014) FigTree. 1.4 ed.

Rambaut A, Lam TT, Max Carvalho L, et al. (2016) Exploring the temporal structure of heterochronous sequences using TempEst (formerly Path-O-Gen). *Virus Evol* 2: vew007.

Reddy EA, Shaw AV and Crump JA. (2010) Community-acquired bloodstream infections in Africa: a systematic review and meta-analysis. *Lancet Infect Dis* 10: 417-432.

Ribeiro TG, Novais A, Peixe L, et al. (2016) Atypical epidemiology of CTX-M-15 among Enterobacteriaceae from a high diversity of non-clinical niches in Angola. *J Antimicrob Chemother* 71: 1169-1173.

Richter SS, Heilmann KP, Dohrn CL, et al. (2009) Changing epidemiology of antimicrobial-resistant Streptococcus pneumoniae in the United States, 2004-2005. *Clin Infect Dis* 48: e23-33.

Roy S, Gaind R, Chellani H, et al. (2013) Neonatal septicaemia caused by diverse clones of Klebsiella pneumoniae & Escherichia coli harbouring blaCTX-M-15. *Indian J Med Res* 137: 791-799.

Rudan I, Boschi-Pinto C, Biloglav Z, et al. (2008) Epidemiology and etiology of childhood pneumonia. *Bull World Health Organ* 86: 408-416.

Rupp ME and Fey PD. (2003) Extended spectrum beta-lactamase (ESBL)-producing Enterobacteriaceae: considerations for diagnosis, prevention and drug treatment. *Drugs* 63: 353-365.

Russo TA and Johnson JR. (2003) Medical and economic impact of extraintestinal infections due to Escherichia coli: focus on an increasingly important endemic problem. *Microbes Infect* 5: 449-456.

Schwarz S, Kehrenberg C, Doublet B, et al. (2004) Molecular basis of bacterial resistance to chloramphenicol and florfenicol. *FEMS Microbiol Rev* 28: 519-542.

Scott JA, Berkley JA, Mwangi I, et al. (2011) Relation between falciparum malaria and bacteraemia in Kenyan children: a population-based, case-control study and a longitudinal study. *Lancet* 378: 1316-1323.

Seale AC, Blencowe H, Manu AA, et al. (2014) Estimates of possible severe bacterial infection in neonates in sub-Saharan Africa, south Asia, and Latin America for 2012: a systematic review and meta-analysis. *Lancet Infect Dis* 14: 731-741.

Seemann T. (2014) Prokka: rapid prokaryotic genome annotation. *Bioinformatics* 30: 2068-2069.

Shah AA, Hasan, F., Ahmed, S., and Hameed, A. (2004) Extended-Spectrum Beta-Lactamases (ESBLs): Characterization, Epidemiology and Detection. *Critical Reviews in Microbiology* 30: 25-32.

Shittu AO and Lin J. (2006) Antimicrobial susceptibility patterns and characterization of clinical isolates of Staphylococcus aureus in KwaZulu-Natal province, South Africa. *BMC Infect Dis* 6: 125.

Smith AM, Govender N, Keddy KH, et al. (2010) Quinolone-resistant Salmonella Typhi in South Africa, 2003-2007. *Epidemiol Infect* 138: 86-90.

Smith CL and Powell KR. (2000) Review of the sulfonamides and trimethoprim. *Pediatr Rev* 21: 368-371.

Snipen L, Almoy T and Ussery DW. (2009) Microbial comparative pan-genomics using binomial mixture models. *BMC Genomics* 10: 385.

Sonda T, Kumburu H, van Zwetselaar M, et al. (2016) Meta-analysis of proportion estimates of Extended-Spectrum-Beta-Lactamase-producing Enterobacteriaceae in East Africa hospitals. *Antimicrob Resist Infect Control* 5: 18.

Spicknall IH, Foxman B, Marrs CF, et al. (2013) A modeling framework for the evolution and spread of antibiotic resistance: literature review and model categorization. *Am J Epidemiol* 178: 508-520.

Stamatakis A. (2014) RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30: 1312-1313.

Sturenburg E, Mack D,. (2003) Extended-spectrum b-lactamases: implications for the clinical microbiology laboratory, therapy, and infection control. *Journal of Infection* 47: 273-295.

Tansarli GS, Athanasiou S and Falagas ME. (2013) Evaluation of antimicrobial susceptibility of Enterobacteriaceae causing urinary tract infections in Africa. *Antimicrob Agents Chemother* 57: 3628-3639.

Tessema B, Kassu A, Mulu A, et al. (2007) Pridominant isolates of urinary tract pathogens and their antimicrobial susceptiblity patterns in Gondar University Teaching Hospital, nothwest Ethiopia. *Ethiop Med J* 45: 61-67.

Tettelin H, Riley D, Cattuto C, et al. (2008) Comparative genomics: the bacterial pan-genome. *Curr Opin Microbiol* 11: 472-477.

Toni T, Welch D, Strelkowa N, et al. (2009) Approximate Bayesian computation scheme for parameter inference and model selection in dynamical systems. *J R Soc Interface* 6: 187-202.

Touchon M, Hoede C, Tenaillon O, et al. (2009) Organised genome dynamics in the Escherichia coli species results in highly diverse adaptive paths. *PLoS Genet* 5: e1000344.

UNAIDS. (2016) Global AIDS update 2016. Geneva, Switzerland: UNAIDS.

Unicef. (2017) Pneumonia claims the lives of the world's most vulnerable children.

Veltman JA, Bristow CC and Klausner JD. (2014) Meningitis in HIV-positive patients in sub-Saharan Africa: a review. *J Int AIDS Soc* 17: 19184.

Ventola CL. (2015) The antibiotic resistance crisis: part 1: causes and threats. *P T* 40: 277-283.

von Gottberg A, de Gouveia L, Tempia S, et al. (2014) Effects of vaccination on invasive pneumococcal disease in South Africa. *N Engl J Med* 371: 1889-1899.

Vouloumanou EK, Karageorgopoulos DE, Rafailidis PI, et al. (2011) Trimethoprim/sulfametrole: evaluation of the available clinical and pharmacokinetic/pharmacodynamic evidence. *Int J Antimicrob Agents* 38: 197-216.

Walker CL, Rudan I, Liu L, et al. (2013) Global burden of childhood pneumonia and diarrhoea. *Lancet* 381: 1405-1416.

Wall EC, Cartwright K, Scarborough M, et al. (2013) High mortality amongst adolescents and adults with bacterial meningitis in sub-Saharan Africa: an analysis of 715 cases from Malawi. *PLoS One* 8: e69783.

Wall EC, Everett DB, Mukaka M, et al. (2014) Bacterial Meningitis in Malawian Adults, Adolescents, and Children During the Era of Antiretroviral Scale-up and Haemophilus influenzae Type b Vaccination, 2000-2012. *Clin Infect Dis* 58: e137-145.

WHO. (2014) Antimicrobial Resistance Global Report on Surveillance. Geneva: World Health Organization.

WHO. (2016) World Malaria Report 2015. Geneva, Switzerland: World Health Organization.
Wong VK, Baker S, Pickard DJ, et al. (2015) Phylogeographical analysis of the dominant multidrug-resistant H58 clade of Salmonella Typhi identifies inter- and intracontinental transmission events. *Nat Genet* 47: 632-639.
Wood DE and Salzberg SL. (2014) Kraken: ultrafast metagenomic sequence classification using exact alignments. *Genome Biol* 15: R46.
Woodford N, Turton JF and Livermore DM. (2011) Multiresistant Gram-negative bacteria: the role of high-risk clones in the dissemination of antibiotic resistance. *FEMS Microbiol Rev* 35: 736-755.
World Bank. (2017a) GDP National Accounts Data.
World Bank. (2017b) Malawi. *Poverty and Equity Data Portal*. (accessed 2017).
World Bank. (2017c) Malawi. *Data*. (accessed 2017).
Wyres KL, Gorrie C, Edwards DJ, et al. (2015) Extensive Capsule Locus Variation and Large-Scale Genomic Recombination within the Klebsiella pneumoniae Clonal Group 258. *Genome Biol Evol* 7: 1267-1279.
Yong D, Toleman MA, Giske CG, et al. (2009) Characterization of a new metallo-beta-lactamase gene, bla(NDM-1), and a novel erythromycin esterase gene carried on a unique genetic structure in Klebsiella pneumoniae sequence type 14 from India. *Antimicrob Agents Chemother* 53: 5046-5054.
Zankari E, Hasman H, Cosentino S, et al. (2012) Identification of acquired antimicrobial resistance genes. *J Antimicrob Chemother* 67: 2640-2644.
Zerbino DR. (2010) Using the Velvet de novo assembler for short-read sequencing technologies. *Curr Protoc Bioinformatics* Chapter 11: Unit 11 15.
Zerbino DR and Birney E. (2008) Velvet: algorithms for de novo short read assembly using de Bruijn graphs. *Genome Res* 18: 821-829.
Zhang Y and Sievert SM. (2014) Pan-genome analyses identify lineage- and niche-specific markers of evolution and adaptation in Epsilonproteobacteria. *Front Microbiol* 5: 110.
Zimba E, Kinney MV, Kachale F, et al. (2012) Newborn survival in Malawi: a decade of change and future implications. *Health Policy Plan* 27 Suppl 3: iii88-103.